

9	146.2	5.9	740	89	AF304371S2	AF304369 Homo sapi
10	144.4	5.8	34593	94	AF348157	AF348157 Mus muscu
11	133.4	5.4	2690	5	AF215894	AF215894 Drosophi
12	133.4	5.4	115873	64	AC017383	AC017383 Drosophi
13	133.4	5.4	175682	4	AC007417	AC007417 Drosophi
14	133.4	5.4	190574	60	AC007352	AC007352 Drosophi
15	133.4	5.4	261846	5	AE003830	AE003830 Drosophi
16	124	5.0	721	89	AF304371S1	AF304371 Homo sapi
17	84.2	3.4	31201	15	SPAC1D4	269239 S.pombe chr
18	71.8	2.9	46335	12	AC022354	AC022354 Arabidops
19	70	2.8	19158	13	AF188714	AF188714 Emericell
20	67.8	2.7	32329	15	SPBC3D6	AL435846 T7 end of
21	67	2.7	32329	15	SPBC3D6	295620 S.pombe chr
22	65	2.6	143209	13	AP001550	AP001550 Oryza sat
23	58.8	2.4	75289	12	AB023046	AB023046 Arabidops
24	58.4	2.4	3665	15	SCYK079C	228304 S.cerevisia
25	56.8	2.3	23533	1	AE001039	AE001039 Archaeogl
26	53.8	2.2	99241	88	AC079030	AC079030 Homo sapi
27	53.8	2.2	145440	69	AC068119	AC068119 Homo sapi
28	53.8	2.2	181289	73	AC068797	AC068797 Homo sapi
29	52.4	2.1	12026	1	AE005109	AE005109 Halobacte
30	51.8	2.1	937	53	CNS06JSP	AL401999 T7 end of
31	51.8	2.1	931	53	CNS06HXX	AL339595 T3 end of
32	51.8	2.1	11836	1	AE000936	AE000936 Methanoba
33	51.4	2.1	729	97	HSY18896	Y18896 Homo sapien
34	51.4	2.1	982	97	HSY18894	Y18894 Homo sapien
35	51.4	2.1	1086	91	HOSAI1573	Y17573 Homo sapien
36	51	2.1	3215	59	HS2VP16B	M57289 Herpes simp
37	50.8	2.0	33779	3	SCG3D3	AL096822 Streptomy
38	50.6	2.0	14189	93	HSPLECTIN	254367 H.sapiens g
39	50.6	2.0	14800	97	HSU53204	U53204 Human plect
40	50.6	2.0	22693	93	HSPEC153	U63610 Human plect
41	49.8	2.0	105795	3	SYCCPNC	D64001 Synchocyst
42	49.2	2.0	226013	74	AC073580	AC073580 Mus muscu
43	48.6	2.0	92858	69	AC024959	AC024959 Homo sapi
44	48.6	2.0	39150	3	MTCY21B4	Z80108 Mycobacteri
45	48.2	1.9	171	59	HS4IR3NA	K00106 epstein-bar

ALIGNMENTS

RESULT	1	AF304370	2908 bp	mRNA	PRI	23-FEB-2001
LOCUS	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
DEFINITION	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
ACCESSION	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
VERSION	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
KEYWORDS	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
SOURCE	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
ORGANISM	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
REFERENCE	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
AUTHORS	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
TITLE	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
JOURNAL	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
PUBMED	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
REFERENCE	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	
AUTHORS	AF304370	2908 bp	mRNA	PRI	23-FEB-2001	

9	146.2	5.9	740	89	AF304371S2	AF304369 Homo sapi
10	144.4	5.8	34593	94	AF348157	AF348157 Mus muscu
11	133.4	5.4	2690	5	AF215894	AF215894 Drosophi
12	133.4	5.4	115873	64	AC017383	AC017383 Drosophi
13	133.4	5.4	175682	4	AC007417	AC007417 Drosophi
14	133.4	5.4	190574	60	AC007352	AC007352 Drosophi
15	133.4	5.4	261846	5	AE003830	AE003830 Drosophi
16	124	5.0	721	89	AF304371S1	AF304371 Homo sapi
17	84.2	3.4	31201	15	SPAC1D4	269239 S.pombe chr
18	71.8	2.9	46335	12	AC022354	AC022354 Arabidops
19	70	2.8	19158	13	AF188714	AF188714 Emericell
20	67.8	2.7	32329	15	SPBC3D6	AL435846 T7 end of
21	67	2.7	32329	15	SPBC3D6	295620 S.pombe chr
22	65	2.6	143209	13	AP001550	AP001550 Oryza sat
23	58.8	2.4	75289	12	AB023046	AB023046 Arabidops
24	58.4	2.4	3665	15	SCYK079C	228304 S.cerevisia
25	56.8	2.3	23533	1	AE001039	AE001039 Archaeogl
26	53.8	2.2	99241	88	AC079030	AC079030 Homo sapi
27	53.8	2.2	145440	69	AC068119	AC068119 Homo sapi
28	53.8	2.2	181289	73	AC068797	AC068797 Homo sapi
29	52.4	2.1	12026	1	AE005109	AE005109 Halobacte
30	51.8	2.1	937	53	CNS06JSP	AL401999 T7 end of
31	51.8	2.1	931	53	CNS06HXX	AL339595 T3 end of
32	51.8	2.1	11836	1	AE000936	AE000936 Methanoba
33	51.4	2.1	729	97	HSY18896	Y18896 Homo sapien
34	51.4	2.1	982	97	HSY18894	Y18894 Homo sapien
35	51.4	2.1	1086	91	HOSAI1573	Y17573 Homo sapien
36	51	2.1	3215	59	HS2VP16B	M57289 Herpes simp
37	50.8	2.0	33779	3	SCG3D3	AL096822 Streptomy
38	50.6	2.0	14189	93	HSPLECTIN	254367 H.sapiens g
39	50.6	2.0	14800	97	HSU53204	U53204 Human plect
40	50.6	2.0	22693	93	HSPEC153	U63610 Human plect
41	49.8	2.0	105795	3	SYCCPNC	D64001 Synchocyst
42	49.2	2.0	226013	74	AC073580	AC073580 Mus muscu
43	48.6	2.0	92858	69	AC024959	AC024959 Homo sapi
44	48.6	2.0	39150	3	MTCY21B4	Z80108 Mycobacteri
45	48.2	1.9	171	59	HS4IR3NA	K00106 epstein-bar

9	146.2	5.9	740	89	AF304371S2	AF304369 Homo sapi
10	144.4	5.8	34593	94	AF348157	AF348157 Mus muscu
11	133.4	5.4	2690	5	AF215894	AF215894 Drosophi
12	133.4	5.4	115873	64	AC017383	AC017383 Drosophi
13	133.4	5.4	175682	4	AC007417	AC007417 Drosophi
14	133.4	5.4	190574	60	AC007352	AC007352 Drosophi
15	133.4	5.4	261846	5	AE003830	AE003830 Drosophi
16	124	5.0	721	89	AF304371S1	AF304371 Homo sapi
17	84.2	3.4	31201	15	SPAC1D4	269239 S.pombe chr
18	71.8	2.9	46335	12	AC022354	AC022354 Arabidops
19	70	2.8	19158	13	AF188714	AF188714 Emericell
20	67.8	2.7	32329	15	SPBC3D6	AL435846 T7 end of
21	67	2.7	32329	15	SPBC3D6	295620 S.pombe chr
22	65	2.6	143209	13	AP001550	AP001550 Oryza sat
23	58.8	2.4	75289	12	AB023046	AB023046 Arabidops
24	58.4	2.4	3665	15	SCYK079C	228304 S.cerevisia
25	56.8	2.3	23533	1	AE001039	AE001039 Archaeogl
26	53.8	2.2	99241	88	AC079030	AC079030 Homo sapi
27	53.8	2.2	145440	69	AC068119	AC068119 Homo sapi
28	53.8	2.2	181289	73	AC068797	AC068797 Homo sapi
29	52.4	2.1	12026	1	AE005109	AE005109 Halobacte
30	51.8	2.1	937	53	CNS06JSP	AL401999 T7 end of
31	51.8	2.1	931	53	CNS06HXX	AL339595 T3 end of
32	51.8	2.1	11836	1	AE000936	AE000936 Methanoba
33	51.4	2.1	729	97	HSY18896	Y18896 Homo sapien
34	51.4	2.1	982	97	HSY18894	Y18894 Homo sapien
35	51.4	2.1	1086	91	HOSAI1573	Y17573 Homo sapien
36	51	2.1	3215	59	HS2VP16B	M57289 Herpes simp
37	50.8	2.0	33779	3	SCG3D3	AL096822 Streptomy
38	50.6	2.0	14189	93	HSPLECTIN	254367 H.sapiens g
39	50.6	2.0	14800	97	HSU53204	U53204 Human plect
40	50.6	2.0	22693	93	HSPEC153	U63610 Human plect
41	49.8	2.0	105795	3	SYCCPNC	D64001 Synchocyst
42	49.2	2.0	226013	74	AC073580	AC073580 Mus muscu
43	48.6	2.0	92858	69	AC024959	AC024959 Homo sapi
44	48.6	2.0	39150	3	MTCY21B4	Z80108 Mycobacteri
45	48.2	1.9	171	59	HS4IR3NA	K00106 epstein-bar

9	146.2	5.9	740	89	AF304371S2	AF304369 Homo sapi
10	144.4	5.8	34593	94	AF348157	AF348157 Mus muscu
11	133.4	5.4	2690	5	AF215894	AF215894 Drosophi
12	133.4	5.4	115873	64	AC017383	AC017383 Drosophi
13	133.4	5.4	175682	4	AC007417	AC007417 Drosophi
14	133.4	5.4	190574	60	AC007352	AC007352 Drosophi
15	133.4	5.4	261846	5	AE003830	AE003830 Drosophi
16	124	5.0	721	89	AF304371S1	AF304371 Homo sapi
17	84.2	3.4	31201	15	SPAC1D4	269239 S.pombe chr
18	71.8	2.9	46335	12	AC022354	AC022354 Arabidops
19	70	2.8	19158	13	AF188714	AF188714 Emericell
20	67.8	2.7	32329	15	SPBC3D6	AL435846 T7 end of
21	67	2.7	32329	15	SPBC3D6	295620 S.pombe chr
22	65	2.6	143209	13	AP001550	AP001550 Oryza sat
23	58.8	2.4	75289	12	AB023046	AB023046 Arabidops
24	58.4	2.4	3665	15	SCYK079C	228304 S.cerevisia
25	56.8	2.3	23533	1	AE001039	AE001039 Archaeogl
26	53.8	2.2	99241	88	AC079030	AC079030 Homo sapi
27	53.8	2.2	145440	69	AC068119	AC068119 Homo sapi
28	53.8	2.2	181289	73	AC068797	AC068797 Homo sapi
29	52.4	2.1	12026	1	AE005109	AE005109 Halobacte
30	51.8	2.1	937	53	CNS06JSP	AL401999 T7 end of
31	51.8	2.1	931	53	CNS06HXX	AL339595 T3 end of
32	51.8	2.1	11836	1	AE000936	AE000936 Methanoba
33	51.4	2.1	729	97	HSY18896	Y18896 Homo sapien
34	51.4	2.1	982	97	HSY18894	Y18894 Homo sapien
35	51.4	2.1	1086	91	HOSAI1573	Y17573 Homo sapien
36	51	2.1	3215	59	HS2VP16B	M57289 Herpes simp
37	50.8	2.0	33779	3	SCG3D3	AL096822 Streptomy
38	50.6	2.0	14189	93	HSPLECTIN	254367 H.sapiens g
39	50.6	2.0	14800	97	HSU53204	U53204 Human plect
40	50.6	2.0	22693	93	HSPEC153	U63610 Human plect
41	49.8	2.0	105795	3	SYCCPNC	D64001 Synchocyst
42	49.2	2.0	226013	74	AC073580	AC073580 Mus muscu
43	48.6	2.0	92858	69	AC024959	AC024959 Homo sapi
44	48.6	2.0	39150	3	MTCY21B4	Z80108 Mycobacteri
45	48.2	1.9	171	59	HS4IR3NA	K00106 epstein-bar

QY 481 atagagctgctgtgctggcccccactctgcccagatacgagatgaaacatlgacgt 540
 DB 481 ATGAGAGCTGCTGTGCTGGCCCCACTCTGCCAGATACGAGATGAACCATGTGACAGT 540
 QY 541 taccagatccccaatacacaagtgaacagagaggggaaagcaccaacatlgcagagttca 600
 DB 541 TACCAGATCCCCATACACAGTGAACAGAGAGGGGAAAGCACCAACCATGTGACAGTCA 600
 QY 601 gaaagcctcgaagcagctcaagtcagtcagagcgtctcagactccagtcgaaatgaaat 660
 DB 601 GAAAGGCTCTCAGAGGAGCTCAGTCCAGAGGAGTCTCAGACTCCGAGTCGAATGAAT 660
 QY 661 gagccacccctccacagtggttagccagagaaaggggtcagagactctccctgttc 720
 DB 661 GAGCCACACCTTCCACATGgtttagccagagaaaggggtcagagactctccctgttc 720
 QY 721 gtagcttcacatctgtaagcttcaacttaagagagaaactctctgtctcacaagcaag 780
 DB 721 GTAGCTTCACTGTGAACCTTCACTTAAGAGAGGAACCTTGTGTCTCAACCAAG 780
 QY 781 gagatggcctcccaagtctggagacagctgcacatccatcatgtctgtctcaagac 840
 DB 781 GAGATGGCCTCCCAAGTGGGACAGCTCCATCCCTCATGTCTGTCTCAAGAC 840
 QY 841 gggaagaacatcacatcagaaggaagagagatlttgcttgaaagactgtgtactctca 900
 DB 841 GGGAAACATCACTCAATGAAGAGAGAGATTTGTGTAAGACTGTGTACTCTCTCA 900
 QY 901 gatcctgtgtctgtctgt 960
 DB 901 GATCCTGTGTCTGT 960
 QY 961 tctggaagatgcacactctcagagtgacaaaggaagagcagatgtccctgtgtgtgt 1020
 DB 961 TGTGAGATGTCACCTTTCAGAGGTACCAAGAGAAAGCAGATGCCCTGTGCTGTGTGT 1020
 QY 1021 gttcaacatggccccagacatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
 DB 1021 GTTCAACATGGCCCCAGACATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
 QY 1081 ttgtgacctgaacccaacactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140
 DB 1081 TTTGGGCTGTGACCCACACACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
 QY 1141 cgcagcacaagatltcaaaacccaagctcaactcaccacccagagacatcttccctgtctc 1200
 DB 1141 CGCAGCCACAAGATTCAAAACCCAGCTCAACCTCATCCACCCGACATCTTCCCTGTCTC 1200
 QY 1201 accagttccgctgttaagaagagagggcccaacctcaagtgtgtgtgtgtgtgtgtgt 1260
 DB 1201 ACCAGTTTCCGCTGTAAAGAGAGGGCCCCACCTTCAGTGTCTCCATGTGTAGGGTGA 1260
 QY 1261 tgcctcctcaagatcacagctccgtccagagagagtgagagagatgtccatttact 1320
 DB 1261 TGCCTCTCAAGTACCAAGTCCGTCCTCAAGAGAGAGTGGCAAGAGATGCCATTATTACT 1320
 QY 1321 tgcacatcttgagaaatcatatgttgagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1380
 DB 1321 TGCATCTCTGAGGAATCATAGTTGAGGGGCTGTGAGCTTCCCAACTTCCAGAGAGCGTG 1380
 QY 1381 cagagatcagagagagtgctgcagagagccagcccaagcagagagaaagaaatcaatc 1440
 DB 1381 CAGGAGTACAGAGAGAGTGCAGAGAGAGCGCCAGCCCGCAGAGAGAGAAAGAGTCAAGT 1440
 QY 1441 ccaagaatacatcttctctggaacagaggtctgcattcccaataagaattgaagtgtcatt 1500
 DB 1441 CCAGAAATCATCTTCTCTGGAACAGAGGTGTGCATCTCCGATGAAGATTCGAATGTCACT 1500
 QY 1501 gccacaactgttcaacataagccccagacagctctgtctactgtgtgtgtgtgtgtgtgt 1560
 DB 1501 GCCACACTGTGTCAACATGAAGCCCGACAGCTGTCTGTGTGTGTGTGTGTGTGTGTGTGT 1560

QY 1561 ttgtggcaagctgtgtccgctcattaccgagacccaggttgagacaggtgtccgtggcaacctgtgt 1620
 DB 1561 TTTGGGCAAGCTGTGCGGTATTAACGAGACAGGTGTGACAGGGTCTGGGGCACTTGCT 1620
 QY 1621 gctgt 1680
 DB 1621 GCTGT 1680
 QY 1681 cagagagaaagcgcctgt 1740
 DB 1681 CAGAGAGAAAGCGGCTGT 1740
 QY 1741 cccaacaaactcaaaagcctgt 1800
 DB 1741 CCCAACCACTTAAGACCTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1800
 QY 1801 caccatcagatgtatctctgt 1860
 DB 1801 CACATCAGATGATTCCTGT 1860
 QY 1861 gtggaagaatgtatcagatgt 1920
 DB 1861 GTGGAAGAATGTATCAGT 1920
 QY 1921 ctgt 1980
 DB 1921 CTGT 1980
 QY 1981 gt 2040
 DB 1981 GT 2040
 QY 2041 accctcctgtatcaatgaagacacccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2100
 DB 2041 ACCCTCTGTATCAATGAACCCACCTGTGAAGATGTGTGTGAAGAGAGAACAGTGAAG 2100
 QY 2101 acaacagacacaagctcccaagccatcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2160
 DB 2101 ACACACAGACAAAGCTCCCAAGCCATCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2160
 QY 2161 atgt 2220
 DB 2161 ATGT 2220
 QY 2221 gagaagtgagagtggt 2280
 DB 2221 GAGAAAGTGGAGT 2280
 QY 2281 cccaagctgtatcccccactgt 2340
 DB 2281 CCCAAGCTGTATCCCTCACTGAAGGCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2340
 QY 2341 cgcagagaaagcggagagctgt 2400
 DB 2341 CGCAGGAGAAACGGAGAGCTGT 2400
 QY 2401 ggcggccttgagagatgt 2460
 DB 2401 GCGGCGCTGAGAGATGGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2460
 QY 2461 aagaagtgtagagccagtgta 2481
 DB 2461 AAGAAAGTCAAGCCCAAGTGA 2481

RESULT 2
 BC001939 2997 bp mRNA PRI 16-MAR-2001
 LOCUS Homo sapiens, putative prostate cancer susceptibility protein,
 DEFINITION clone MGC:4102, mRNA, complete cds.
 ACCESSION BC001939
 VERSION BC001939.1 GI:12804972
 KEYWORDS MGC.

QY 1141 cgcagcacaagattcaaacacagctcaacatccatccaccccgagacatctcccccgtc 1200
 |||||
 Db 1201 cgcagcacaagattcaaacacagctcaacatccatccaccccgagacatctcccccgtc 1260
 QY 1201 accagttccgctgttaagaagagagcccaacccctcaagtggtccatggttcaagggtga 1260
 Db 1261 accagttccgctgttaagaagagagcccaacccctcaagtggtccatggttcaagggtga 1320
 QY 1261 tgcctccctcaagttacagagctccggtcccaagagagagtggaagagagatgcaatttact 1320
 Db 1321 tgcctccctcaagttacagagctccggtcccaagagagagtggaagagagatgcaatttact 1380
 QY 1321 tgcacatccctggagaaatcatagttagttgagagcgttcagcttcccaacttccagaagagcgt 1380
 |||||
 Db 1381 tgcacatccctggagaaatcatagttagttgagagcgttcagcttcccaacttccagaagagcgt 1440
 QY 1381 caggagttacagagagagtggtcagagagcagcccaagcagagagagaaagagagttac 1440
 |||||
 Db 1441 caggagttacagagagagtggtcagagagcagcccaagcagagagagaaagagagttac 1500
 QY 1441 ccagaaatcatcttcccttggaacagaggtctgcatcccgatgaagattccgaattcaggt 1500
 |||||
 Db 1501 ccagaaatcatcttcccttggaacagaggtctgcatcccgatgaagattccgaattcaggt 1560
 QY 1501 gccacactgtgcaacatagagcccgagagcgtctgctacttgagactgtgtgagagcaca 1560
 |||||
 Db 1561 gccacactgtgcaacatagagcccgagagcgtctgctacttgagactgtgtgagagcaca 1620
 QY 1561 ttggagcagctgtgtgcgcatctacagagagcaggtggaagaggtctcctggcagccttggt 1620
 |||||
 Db 1621 ttggagcagctgtgtgcgcatctacagagagcaggtggaagaggtctcctggcagccttggt 1680
 QY 1621 gctgtgtgtgtgtccacactgcaacagagagagagagagagagagagagagagagagag 1680
 |||||
 Db 1681 gctgtgtgtgtgtccacactgcaacagagagagagagagagagagagagagagagagag 1740
 QY 1681 cag 1740
 |||||
 Db 1741 cag 1800
 QY 1741 ccccaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1800
 |||||
 Db 1801 ccccaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1860
 QY 1801 cacaatcagatgatctctgcaaatgcttcaagaagagagagagagagagagagagagagag 1860
 |||||
 Db 1861 cacaatcagatgatctctgcaaatgcttcaagaagagagagagagagagagagagagagag 1920
 QY 1861 gtggaagaatgatgatgctgtgtgtgcaaatgcttcaagaagagagagagagagagagag 1920
 |||||
 Db 1921 gtggaagaatgatgatgctgtgtgtgcaaatgcttcaagaagagagagagagagagagag 1980
 QY 1921 ctgt 1980
 |||||
 Db 1981 ctgt 2040
 QY 1981 gt 2040
 |||||
 Db 2041 gt 2100
 QY 2041 accctccctgatacatgag 2100
 |||||
 Db 2101 accctccctgatacatgag 2160
 QY 2101 aacacacagacacag 2160
 |||||
 Db 2161 aacacacagacacag 2220
 QY 2161 atgt 2220
 |||||
 Db 2221 atgt 2280

QY 2221 gagaagatgt 2280
 |||||
 Db 2281 gagaagatgt 2340
 QY 2281 ccccaagctgtatcccccactgaaagccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2340
 |||||
 Db 2341 ccccaagctgtatcccccactgaaagccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2400
 QY 2341 cgcag 2400
 |||||
 Db 2401 cgcag 2460
 QY 2401 ggcggcctgtgagagatgt 2460
 |||||
 Db 2461 ggcggcctgtgagagatgt 2520
 QY 2461 aagaagttcag 2481
 |||||
 Db 2521 aagaagttcag 2541

RESULT 3
 BC004158
 LOCUS BC004158 3006 bp mRNA PRI 16-MAR-2001
 DEFINITION Homo sapiens, putative prostate cancer susceptibility protein,
 clone MGC:2441, mRNA, complete cds.
 ACCESSION BC004158
 VERSION BC004158.1 GI:13278770
 KEYWORDS MGC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 3006)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.L.Strausberg@nih.gov
 Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadan@systemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
 Greene, Mark Ketterman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: ITAL Plate: 2 Row: C Column: 13.
 Location/Qualifiers
 1. 3006
 /organism="Homo sapiens"
 /db_xref="LocusID:60528"
 /db_xref="taxon:9606"
 /clone="MGC:2441 IMAGE:2820640"
 /tissue_type="Lung, small cell carcinoma"
 /tissue_lib="NIH_MGC_7"
 /lab_host="DH10B-R"
 /note="Vector: pORB7"
 70..2550
 /codon_start=1
 /product="putative prostate cancer susceptibility protein"
 /protein_id="AAH04158.1"
 /db_xref="GI:13278771"
 /translation="MMAALCSILRSAGRTMSQGRITSOAPARRRPRKDPRLRLTRE
 KRPSGCSGGNTYLYLVVAGSRDSGAALVVFSEFNRYLFNCGSGVRLMGEHLKY

FEATURES

source
 CDS

[illegible]

Db	1930	GTGAAAGATTGATGACAGTTGGCTGTGGCAACATGATTGGATTGGAGAGTTTCAGACCTGT	1989
Qy	1921	ctggctgcggcaactcgcgaagcatgcgttctgtgcctctgtgcctcgtgcacactctcgtcctgaa	1980
Db	1990	CTGGTGGGGACATCGCAAGCATGTGGTTGGCTGTGGCCTGTGTACACCTCTGGCTGGGAAA	2049
Qy	1981	gtgtgtcattccggggagaccacagccctcgcgaagctcgtgctcggatbtgggaaagaatgcc	2040
Db	2050	GTGGTCTATTCCGGGGACACACATGCCCTCGAGAGCCTGTGTGGATGGGAAAGATGCC	2109
Qy	2041	accctccgtatcatgaagcaccctctgaaagatggtlttgaaagagaaacagltgaaag	2100
Db	2110	ACCTCTCTGTATCATATGAAGCACCTCTGGAAAGATGTTTGGAGAGCAAGCAGTGGAAAG	2169
Qy	2101	acacacagcaacaagctcccaagacatcaagctcgtgggaatgcgaataaacggaattcat	2160
Db	2170	ACACACAGCAACAGTCCCAAGCCATCACGtgggagatcggatGAACGGAGTTCATT	2229
Qy	2161	atgctgaacaacctctcagccagcgctatgtccaaagtcccccctcttaagccccaacttcagc	2220
Db	2230	ATGCTGAACCACTTCAGCCAGCGCCTATGTGCAAGTGTCCTCTTCAGCCCAACTTCAGC	2289
Qy	2221	gagaaagtctgggaattgcctcttgaccacatgaaggtctgctttggagaatttccaacatg	2280
Db	2290	GAGAAATGTGGAACTGTGCTTTGACCAATGAAGAGTCTGTTTGGAGACTTTTCCAACAATG	2349
Qy	2281	cccaagctctgatccccaactgaaagaccctgttctgtcgtgcacatcgaagagaatggaag	2340
Db	2350	CCCAAGCTGATTCGCCCACTGAAGAGCCGTGTTTGTGGGAGCATGTGAGAGATGGAGAG	2409
Qy	2341	cgcacggagaagaagcgggaactcgcgcaggtgtgcggcgcgccctcctgtccaaggaagcttgca	2400
Db	2410	CGCAGGGAGAGCGGGAGCTGCGCAGGtgcCGGGCGGCTCTCTTCCAGGGAGCTTGCA	2469
Qy	2401	ggcgcccttgaaagatctgggaagcctctcagcagaagcggccacacagagaagaccagaagc	2460
Db	2470	GGCGGCTGTGAGATGGGAGGCTTCAGAGAAAGGGGGCCACACAGAGAGGCCACAGGCC	2529
Qy	2461	aagaagctcagaagcccaagtga 2481	
Db	2530	AAGAAGTCAAGACCCACGTA 2550	

RESULT	4
LOCUS	AF308698
DEFINITION	Pan troglodytes ELAC2 mRNA, complete cds.
ACCESSION	AF308698
VERSION	AF308698.1 GI:10946496
KEYWORDS	
SOURCE	Chimpanzee.
ORGANISM	Pan troglodytes
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.
AUTHORS	1 (bases 1 to 2908) Tevligian,S.V., Sismard,J., Teng,D.H.F., Abtlan,V., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C., Gaffari,S., Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Korl,E.N., Lalit,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,J., Schroeder,M., Smith,B., Snyder,S.C., Swedlund,B., Swensen,S., Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhansen,S., Komenus,J., and Cannon-Albright,L.A.
TITLE	A candidate prostate cancer susceptibility gene at chromosome 17p
JOURNAL	Nat. Genet. 27 (2), 172-180 (2001)
PUBMED	11175785
REFERENCE	2 (bases 1 to 2908) Tevligian,S.V., Sismard,J., Teng,D.H.F., Baumgard,M., Dayananth,P., Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C., Gaffari,S., Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Korl,E.N., Lalit,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,

[illegible]

OY	361	ttaaagtgaatgatcttactttaaaggaaaccgggcttccaagtgtgtacttctg	420
Db	361	TTAAAGTGAAGAATCTTAACTTTAAAGGAAACCGGGCTTCCAAAGTGTACTTCTTGGA	420
OY	421	cctccacaactcgggaaaaatactcctcgaagaacataaaataatttctcgtccattgaaaga	480
Db	421	CCTCCACAACCTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGTGTCATTGAAGA	480
OY	481	atagaactcgtcgtcgcggcccaactcgcgcccaatacagatgaaacatgacagt	540
Db	481	ATAGAAGCTGGCTGGCGGGCCCACTCGCCCGAAGATAGAGAGATGAACAATGACAGATT	540
OY	541	taccaagatcccatacacacagatgaaacagagagaggaaagaccacaacatgycagatcca	600
Db	541	TACCAAGATCCCACAAACACAGTGAACAGAGAGGGGAAACACCAACCATGTGGAGAGTCCA	600
OY	601	gaaggagctctcaacagagctcaagtcacagagagatcttaagatccagatcgatgaaataat	660
Db	601	GAAGAGCTCTCAGCAGGCTCAAGTCCAGACGATCTTCAAGACTCCGAGTCAAAATGAATAAT	660
OY	661	gagcacacacttcacacatagttgttaacagagaagaagggttaaggagactcttcctgtctc	720
Db	661	GAGCACACACTTCCACATGSGTGTAAOCAGAGAGAGGGGTGAGGAGACTCTTCCTGGTCC	720
OY	721	gtagcttcatctgttaagcttcaactaaagagagaaactcttggtgtcacaagaacaag	780
Db	721	GTACTTTCATCTGTAACTGCTTCACTTAAAGAGAGAAACTTGTGTGTCGTCAAAACAAG	780
OY	781	gagtgtaggccccccagttgggaagctgcacatgcgtcccatattgtctgtctgtaagagac	840
Db	781	GAGATGGGCTCCCAAGTTGGGACAGCTGCATCGCCCATATTGCTGTCTCAAGGAC	840
OY	841	gggaaagaacatcaactcaatgaagaagaagagatttgcctgaagactgtgtaactccaca	900
Db	841	GGGAAAACCATCACTCACTGAAGAGAAGAGATTTTGGCTGAAGAGCTGTACTCTCTCA	900
OY	901	gacctgtgtcgtcttctgtgtgtgtaagaatgtccagatgaaagcttcaatcaaccalc	960
Db	901	GATCCTGTGTGCTGCTTTGHTGTGTGAATGTCAGATGAAGGCTCATTCACACCAATC	960
OY	961	tgtgaagaatgcacacttccaaagtatccaaagaaagcagatgtccccgtgtgcctgtgtg	1020
Db	961	TGTGAGATGTCACCTTTGACAGAGTACCAAGGAAAGCAGATGCCCGTGCCTTTGSTG	1020
OY	1021	gttcaatgccccagaactgtgcttgtgtggaagaaggtaccagcagttgatatggaagag	1080
Db	1021	GTTCAATGGCCCCAGATCTGTGCTTGTGTGAGAGGAGGTACACAGATGGATGGAAGAGG	1080
OY	1081	tttggagctgaacacccagcaactcgtctccctgaatgaaagacttgcctcagttcaacaact	1140
Db	1081	TTTGGGCTTGACACCCAGCACTTGTGTCGTGAATGAACATGTGGCTCACTCAACACTT	1140
OY	1141	cgcagacacaagatccaaccacagctcaactcaccatccaccggagacactctcccccgtc	1200
Db	1141	CGCAGCCACAAGATTCAAAACCCAGCTCAACCTCATCCACCAGCAATCTTCCCCCTGCTC	1200
OY	1201	accagtttcgcgtgtgaagaagagagccccacacctgaagtgtcccatgtttcaagvtgaa	1260
Db	1201	ACCAAGTTTCCCTGTAAAGAGGAGGCCCCACACTTGAAGTGTCCCATGGTTCAAGGTGAA	1260
OY	1261	tgctccctcaagtacacagctccgttcccagagagagatgtgcaagagatgycatcttact	1320
Db	1261	TGCTCCTCTCAAGTACACAGTCCGTCCACGAGAGAGATGGCAAGGATGTCCATTATTAAT	1320
OY	1321	tgcgaatccgaagaaatcatagtttgaagcgctgtcagcttcccaacttccagcaagacgtg	1380
Db	1321	TGCATCTCTAGGAATTCATAAATTTGAGGCGCTGCACACTTCCCAACTTCACAGAGAGTGTG	1380
OY	1381	cagagatgaagagagagatgtgcagagagcgccagcccccagcagaaaaaagaatcagtaac	1440
Db	1381	CAGAGATGACAGAGAGAGATGTGCGAGAGCGGCCACGCCCAAGCAGAAAAAGAGTACAGTAC	1440

QY	1441	ccagaatcatcttcccttggaaagggctgcgaatccgaataaattcgaaatctg	1500
Db	1441	CCAGAAATCATCTTCTTTGGAAACAGGGTGGCCATCCCATATAATTCGAAATCTCAGT	1500
QY	1501	gccacactctgtcaacataaagccccgaacagctctctctactctagactgtgtgagggca	1560
Db	1501	GCACACATTGTCAACATTAAGCCCCGACAGCTCTCTACTCTAGCATGTGTGTAGGGCAGG	1560
QY	1561	tttgggcagctgtgtcgctcatctacgagagccagvtgaaaggtccctggtgcaacctgtct	1620
Db	1561	TTTTGGCGAGCTGTCCCTCTCAATTACGGAGACCAAGTGGACAGGGTCTCTGGCACCTGTGCT	1620
QY	1621	gctgtgtttgtgtcccaactgtcaacgagatcaccaacaggggtctgcagaatctctgtg	1680
Db	1621	GCTGTGTTTGTGTGCCACTGTGCACGACGAGATACACACACGGGCTTCTAATAATCTTGGTG	1680
QY	1681	cagagagaagcgcgctctgtgcatctttgggaaagcgcttcacaccttgcvtgtgtgtgc	1740
Db	1681	CAGAGAGAAGCAGCGCTTGGCATCTTTGGGAAACCCCTTTCACCTTTTGCTGTGTGTGCC	1740
QY	1741	cccaaccaggtccaaagctgtgtcccaagataccacaaccagttgcagaagagttcctgac	1800
Db	1741	CCCAACACAGCTCAAAAGCTGTGCTCCACGACATACACACAGGCGCAGAGGTTCTTGAC	1800
QY	1801	cacatcagiatgatcttccctgcacaaatgtcccttcagagaaagggcgctgaatctcagttcga	1860
Db	1801	CACATCAGTATGATTCCTGGCAATAGTCCTTACAGAAAGGGGCTGTGACATGCCAGTCTCTCA	1860
QY	1861	gtggaagaatgatcatcagttctgcgtgtgtgcgaacatgtgtatttggaaagagttccaagctgt	1920
Db	1861	GTGGAAGAATGATGATTCGTGTGCTGTGGCGAACAATGTGATTTGGAAAGATTTCAGACCTGT	1920
QY	1921	ctgtgtgcggcaactcgaagcatgctgttggctgtgtgcgtgtgtgcacacctctgcgtcgaa	1980
Db	1921	CTGTGTGCGGCACTCAGACATGCGTTTGGCTGTGGCCTGTGTGCACACCTGTGTGCTGGAA	1980
QY	1981	gtgtgtcatttccggygaacacatgtccctgcgagagctctgtctcggatlggygaagaatgccc	2040
Db	1981	GTGTGTCTATTTCGGGAGACACATGACCCTCGAGAGGCTCTGTGCGATGGGAAATATGCC	2040
QY	2041	accctccctgataacttgaagccacccctgggaagatgtgttggaaagaaagcagttgaag	2100
Db	2041	ACCTCTCGATACATVGAAGCCACCTCTGGAAAGCGTGTGGAAAGAAAGCATGTGAAAG	2100
QY	2101	aacacagcaacaagctcccaagccatcgaagtgtyggatgtcggatgaacgcgagatcatt	2160
Db	2101	ACACACACACAAAGCTCCCAAGCATACGATGATGGGATGTGGATTAACGGGAGTTCAATT	2160
QY	2161	atgtctgaaccaactcaagcagcgctabgcgaagttccccccttcagcccccaactcagc	2220
Db	2161	ATGTCTGAACCACTTCACGACGCGCTATGTGCACAAAGTCCCTCTTAGCGCCCAACTTCAC	2220
QY	2221	gagaaagtgggaagttgtcctttgaacacatgaagatctgttttggagacttccaacaatg	2280
Db	2221	GAGAAAGTGGGAGTTGTCCCTTTGACACATGABAGGTGTGCTTTGGAGACTTTGCCAACAATG	2280
QY	2281	cccaagctgtatcccccacatgaagaccctgtttgtctgcgacatctcgagagagaatggaag	2340
Db	2281	CCCAAGCTGATTTCCCCACATGAAGGCCCTTTGTCTGGGACATGTGAGAGATGTGAGAG	2340
QY	2341	cgccagggagaagcgggagctgtgcggcagggtgtgcggcgccctctgtccaaggagctgca	2400
Db	2341	CGCAGGGAGAAAGCGGGAGCTGCGGACAGGTGCGGGCGCCCTCTCTTCACAGGGAGCTTGCA	2400
QY	2401	ggcgccctggagagaatggggagcctcaagcaagaacgggccccacagagagagccacagcc	2460
Db	2401	GGCGCCCTGGAGAGATGGGGAGCCTCAAGCAAGAACGGGCCCAACACAGAGAGGACACAGCC	2460
QY	2461	aagaaagtccagagcccaatga 2481	
Db	2461	AAGAAAGTCCAGAGCCCAAGTA 2481	

RESULT	5
AK001392	
LOCUS	AK001392 2976 bp mRNA PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10530 fls, clone NT2RP2000985.
ACCESSION	AK001392
KEYWORDS	AK001392.1 GI:7022621
SOURCE	Oligo capping, fls (full insert sequence).
ORGANISM	Homo sapiens tetratocarcinoma cell_line:NT2 cDNA to mRNA, clone_1fb:NT2RP2 clone:NT2RP2000985. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (sites)
REFERENCE	Isoqal,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shitatori,A., Sudo,H., Wagatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuo,Y., Nhomiya,K. and Iwayanagi,T. NEO human cDNA sequencing project Unpublished (2000) 2 (bases 1 to 2976) Isoqal,T. and Otsuki,T. Direct Submission Submitted (16-FEB-2000) to the DDBJ/EMBL/Genbank databases. Takao Isoqal, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@rri.co.jp, Tel:+81-438-52-3951, Fax:+81-438-52-3952) NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection:: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
COMMENT	Location/Qualifiers 1..2976 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_line="NT2" /cell_type="tetratocarcinoma" /clone_1fb="NT2RP2000985" /note="Cloning vector: PME18FPL3: mRNA from NT2 neuronal precursor cells after 2 weeks retinoic acid (Ra) induction." 75..2639 /note="unnamed protein product" /codon_start=1 /protein_id="BAA91666.1" /db_xref="gi:7022622"
FEATURES	/translation="MMALCSILBRASAGRTMSQGTISDAPARRRRPKDPRLRHRTRE KRSGSGGCGPNVYTLQVAAGSRDSOGALVFSEFNNTLNGEGVGQRMQDEKLKV ARLDNIETLRHMWSNVGSIMLTITETGTPKCVLGGPOLLELAIKTFSPGLKVV IELMRPHSAPEYEDEDTMTVYOIPIHSSORRKGQPOMSPERPRLSLPERSSDSEN ENEBELDHGVSORRGVDSSLVAFICFLHLKRNPFYLAKMGDLPGTAIAPIITA AVKGGKSTHGREGLEAELECTPPDGCAFPVYGVSFGPTOPICENATFOROYGGKAD APVALVMAASIVADYSRYQOMMERGPOTQHLYLNENCASYNNLSHSKTOTDNILI HPDFPLITSFRCKEGPLSPVMQGEQLTYOLPRREWRODIITCNPEFTVEAL QLPFGQSVODEYRSADQPAAPARKSQEPPIIFLGGSALIPMKIRNVSATLVNPIS TSLILDCEGFFGOLCRHYGDQDVRYGLTAAVNSHIHADHRGELSIILODRALALA SIGLPDLHVAVAPNOUKAMLOOQYNYCCOEVLNISMADCOEGAEISSPARELLIT SLKRTGDIEFORCLVRHCNHARGCALVHTSGMKIYVSGDTPGCALVPMGRKATLI IHEATIDGEEZEAVERKTHSTTSOALSIVGMANNPFLTNHFSORYAKVPFSPNPSS KVGAFTDMKVCLEDGFPMKPILPPLKALEGDIIEEMEREREKELQVRALLSLREL AGGLEDDSLSSGTSQRSHRPREPSPEPDIEDPLRLCYCPTHAPVASLALAGRS"
CDS	
BASE COUNT	712 a 807 c 856 g 601 t
ORIGIN	
Query Match	98.7%; Score 2448.2; DB 89; length 2976;
Best Local Similarity	99.7%; Pred. No. 0;

[illegible]

1081 ttgggctacacacccagccttgctctgaatgagaactgtgctcctcaagttccaaactt 1140
1155 ttggggcctacacacccagccttgctctgaatgagaactgtgctcctcaagttccaaactt 1214
1141 cgcagccacaagattcaaacccagctcaacccatccaccccgagacatctccctgtctc 1200
1215 cgcagccacaagattcaaacccagctcaacccatccaccccgagacatctccctgtctc 1274
1201 accagttccgctgtgaagagagggcccccacccagctgagtgccatggttccaggtgaa 1260
1275 ACCAGTTTCGGCTGT--AAAGGAGGGCCCAACCTCAGTGTGCCATGTTCAAGGGTAA 1331
1261 tgcctcctcaagtaacagcccgctcccaagagggagtgagagagagagtgccatttact 1320
1332 TGCTCTCTCAAGTACAGCAGCTCCGTCACAGAGGAGAGTGCAGAGAGATGTCATTATTA 1391
1321 tgcacccctgaagaaatcagttgagagcgctgcaagcttcccaactccagcagagcgctg 1380
1392 TGCATCTCTGAGGAATTCATAGTTGAGGCGCTGACAGCTTCCAAACTTCCAGCAGAGCGTG 1451
1381 cagagatcacagagagagtgagcagagcagcccgcccgagcagaagaaagagtgaglac 1440
1452 CAGGAGTACAGGAGAGTGCAGAGAGCGCCACGCCACAGAGAAAGCAAGTCAAGTAC 1511
1441 ccagaaatcattctccttggaacagagctgcaatcccgatgaagattcgaaatgtcagt 1500
1512 CCAGAAATCATCTTCTTGGAAACAGGCTGTCATCCCGATGGAAGATTCGAAATGTGAGT 1571
1501 gccacactgtcaacalaagcccccgagacgctctgtctactgtgagtgagtgagcaca 1560
1572 GCCACACTTGTCAACATTAAGCCCCGACAGCTCTGTCTACTGTGAGTGTGAGGGGACA 1631
1561 ttggggcagctgtgctgcttcttaagagacagagtgagacaggttccctgggacacctggt 1620
1632 TTTGGGCGAGTGTGCGCTCTTACGAGACAGAGTGAGGAGGCTCTGGGACCCCTGGCT 1691
1621 gctgttcttgctccacacccagcagcagatcacacacccgggctgtgccaatctctgtctg 1680
1692 GCTGTCTTGTGTCACCCAGCGACAGATCACACAGGCGCTTGCACATATCTTCTGCTG 1751
1681 cagagagaacgcgctctgtgcatcttggaagacgcgctcaacctgtctgtgtgtgctc 1740
1752 CAGAGAGAAGCGCCCTGTGGCATCTTGGGAAAGCCGCTTCAACCTTGTGCTGTGCTGCC 1811
1741 cccaacagctcaagcgtgtgctccagcagatcacacacccagtgccagagaggtctgtcac 1800
1812 CCCAACAGCTCAAAAGCCTGTGCTCAGCATCTACAAACAGTGCCAGAGGCTCTGCAC 1871
1801 cacatcagtaatgattcctgtcacaatgctcctcaagaaaggggctgagatctcagctctgca 1860
1872 CACATCAGTATGATTCCTGCGCAATGCTTCAGGAAGGGGCTGAGATCTCCAGTCCGCA 1931
1861 gtygaaagatgtagcagtgctgtgtgcaaatgtgaaattggaagagtttcagacctgt 1920
1932 GTGGAAAGATGTGATCGTGTGTCGCAAAATGTGATTTGGAAGAGTTCAGACCTGT 1991
1921 ctgtgtcgcgcacgtcaagaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1980
1992 GTGGTGGGAGCTGCAAGAGT 2051
1981 gtygtatattccgggagacacacgctcgtgagagctctgttccgagtgaggaaagatgccc 2040
2052 GTGTGTATTTCGGGGACACACATGCTCGTGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGT 2111
2041 accctcctgatacagagccacccctggaagatgtgttggaagaggaagcagtggaag 2100
2112 ACCCTCTGTATCATGATGAAGCCACCTGGAAGATGTTTGAAGAAGCAAGTGTGAAG 2171
2101 acccacaagcacaacgtcccaagcctcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2160
2172 ACACACAGCAACAGTGTCCCAAGCCTATCAGCTGTGGGATGCGGATGAAGCGGAGTTTCATT 2231

QY 2161 atgtgaaccacttcagccagcgtctatgccaaggtccccctcttcagcccccaactcagc 2220
Db 2232 ATGCTGAACCATCTTACGCCAGCGGTATGCAAGAGTCCCTCTTACGCCCACTTACG 2291
QY 2221 gagaagatggagatgtgcttgcacacatgaaggtcgtctgtgtgtgtgtgtgtgtgtgtgt 2280
Db 2292 GAGAACTGGAGATGTGCTTTGACCAATGAGGTGTGCTTGTGAGACTTTCGAACAAATG 2351
QY 2281 cccaagctgattcccccaactgaaagccctgttgcgtgagacatcgagagatggaag 2340
Db 2352 CCCAAGCTGATTTCCCTACACTGAAAGCCCTTGTGCTGACATCAGAGGATGAGAGAG 2411
QY 2341 cgcaggaagaagcggagagctgcgcaggtgctggcgagccctctgttccaggaagctgca 2400
Db 2412 CGCAGGAGAAAGCGGAGAGCTGCGGAGAGTGCGGGCGGCTCTGTCTCAGAGAGCTGCA 2471
QY 2401 ggcgcctggaagatggaagagctcagcagaagcgggcccacagagagagccagagcc 2460
Db 2472 GCGCGCTGGAGGAT--GGGAGCTCAGACAGACGGGCCACACAGAGAGCCACAGGCC 2530
QY 2461 aagaagatcagagccagatga 2481
Db 2531 AAGAGGTCAAGGCCAGTGA 2551

RESULT 6
AF308694
LOCUS AF308694 2893 bp mRNA PRI 27-FEB-2001
DEFINITION Gorilla gorilla ELAC2 (ELAC2) mRNA, complete cds.
ACCESSION AF308694
VERSION AF308694.1 GI:10946488
KEYWORDS
SOURCE
ORGANISM
Gorilla gorilla
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Gorilla.
REFERENCE
1 (bases 1 to 2893)
Tavligian,S.V., Sismard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
Chaffari,S., Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N.,
Lally,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
Neuhansen,S., Rommens,J. and Cannon-Albright,L.A.
A candidate prostate cancer susceptibility gene at chromosome 17p
Nat. Genet. 27 (2), 172-180 (2001)
11175785
TITILE
JOURNAL
PUBMED
2 (bases 1 to 2893)
REFERENCE
Tavligian,S.V., Sismard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
Dumont,M., Farnham,J.M., Frank,D., Frye,C., Chaffari,S.,
Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E.,
Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhansen,S., Rommens,J.
and Cannon-Albright,L.A.
Gorilla gorilla ortholog of human HPC2/ELAC2
Unpublished
TITILE
JOURNAL
REFERENCE
3 (bases 1 to 2893)
AUTHORS
Tavligian,S.V., Sismard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
Dumont,M., Farnham,J.M., Frank,D., Frye,C., Chaffari,S.,
Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E.,
Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhansen,S., Rommens,J.
and Cannon-Albright,L.A.
TITLE
JOURNAL
Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt


```

Db      1621 GCATGTTGTCGCCACCTGCAAGCAGATCACACAGCGGGCTTGTAAATATCTTCCTG 1680
Qy      1681 cagaagagaacgccttgcatcttgggaagcgccttaaccttgcgtggtgtgcc 1740
Db      1681 cagaagagaacgccttgcatcttgggaagcgccttacccttgcgtggtgtgcc 1740
Qy      1741 ccaaccagctcaaacgcttgctccagcagtacaccacacagctccagaggtcttcac 1800
Db      1741 ccaaccagctcaaacgcttgctccagcagtacaccacacagctccagaggtcttcac 1800
Qy      1801 cacatcagatgatcttcctgcacaatgcttcaagaaagggtgagatccagctcgtca 1860
Db      1801 CACATCAGTATGATTCCTGCAAAATGCTTCAGAGAAAGGGCTGAGATCTCCAGCTCGCA 1860
Qy      1861 gtggaagatgatcagctgctgctgtgctgcaaacatgtattggaaaggttccagactgt 1920
Db      1861 GTGGAAGATGATGATTCCTGCAAAATGCTTCAGAGAAAGGGCTGAGATCTCCAGCTCGT 1920
Qy      1921 ctggtgcgacgtgcaagcagatgctgctgtgctgctgctgctgctgctgctgctgctg 1980
Db      1921 CTGCTGGGACATGCAAGCATGCGTTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
Qy      1981 gtggtctatctcgggggaacacatgcccctgcgaggtctctggtccgagatgggaaagtcc 2040
Db      1981 GTGGTCTATTCGGGGGACACCATGCTCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
Qy      2041 acctcctgtatagatgagcagccttggaagatggtttggaaaggaagcagtggaagaa 2100
Db      2041 ACCCTCTGTATGATGAAAGCACCCTTGGAAGATGCTTTGGAAGGAAAGAGTGGAAG 2100
Qy      2101 acacacagcacaagctcccaagcagcagcagcagcagcagcagcagcagcagcagcagc 2160
Db      2101 ACACACAGCACAAGCTCCCAAGCATGAGCGTGGGAGATCGGAGTGAACCGGAGTTTATT 2160
Qy      2161 atgctgacaacattcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2220
Db      2161 ATGCTGAAACCATTCAGCCGCTGATGCAAGGATGCCCTCTTCAGCCCAACTTCAC 2220
Qy      2221 gagaagaatggagatgctgcttgcacacatgaagatgctgcttgcagacattccaaagt 2280
Db      2221 GAGAAATGGAGATGCTGCTTGCACCATGAAAGTGTGCTTTGGAGACTTTCACCAATG 2280
Qy      2281 cccaagctgattcccccacatgaaagccctgttgcctgcagacatcagagagatggaag 2340
Db      2281 CCAAGCTGATTCCTCCACATGAAAGCCCTGTTGCGGCGACATCGAGAGATGAGAGAG 2340
Qy      2341 cgcagaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2400
Db      2341 CGCAGGAGAGAGCGGAGCTGCGGAGTCCGGCGGCTCTCTGTCGGGAGACTGGCA 2400
Qy      2401 ggcgccttggaagatgggagcctcagcagaagagggccacacagaagagggccacagggc 2460
Db      2401 GCGCGCTTGAGAGTGGGAGCCTCAGCAGAAAGGGCCACACAGAGAGGCCACAGGCC 2460
Qy      2461 aagaaggtcagagccagatga 2481
Db      2461 AAGAAGGTCAAGGCCAGTGA 2481

```

```

RESULT 7
AF308696 2712 bp mRNA ROD 27-FEB-2001
LOCUS Mus musculus ELAC2 mRNA, complete cds.
DEFINITION AF308696
VERSION AF308696.2 GI:11992378
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2712)
AUTHORS Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,

```

```

Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
Ghaffari,S., Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N.,
Lally,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
Neuhansen,S., Rommens,J., and Cannon-Albright,L.A.
A candidate prostate cancer susceptibility gene at chromosome 17p
Nat. Genet. 27 (2), 172-180 (2001)
1115785
TITLE
JOURNAL PUBMED
REFERENCE
AUTHORS
Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
Gupte,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E.,
Leavitt,A., Leblanc,G., McArthur-Morrison,J., Peterson,A., Penn,B.,
Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhansen,S., Rommens,J.,
and Cannon-Albright,L.A.
Mouse ortholog of human HPC2/ELAC2
Unpublished
TITLE
JOURNAL
REFERENCE
AUTHORS
Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
Gupte,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E.,
Leavitt,A., Leblanc,G., McArthur-Morrison,J., Peterson,A., Penn,B.,
Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhansen,S., Rommens,J.,
and Cannon-Albright,L.A.
Direct Submission
Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
Lake City, UT 84103, USA
4 (bases 1 to 2712)
TITLE
JOURNAL
REFERENCE
AUTHORS
Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
Gupte,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E.,
Leavitt,A., Leblanc,G., McArthur-Morrison,J., Peterson,A., Penn,B.,
Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhansen,S., Rommens,J.,
and Cannon-Albright,L.A.
Direct Submission
Submitted (26-DEC-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
Lake City, UT 84103, USA
Sequence update by submitter
On Dec 26, 2000 this sequence version replaced gi:10946492.
location/Qualifiers
1.2712
/organism="Mus musculus"
/db_xref="taxon:10090"
/dev_stage="fetus"
15.2510
/note="similar to Homo sapiens HPC2/ELAC2"
/codon_start=1
/product="ELAC2"
/protein_id="AAG24918.2"
/db_xref="GI:11992379"
/translation="MVALRSILAPLGLRTMSQGSARPRPSKDLRLHRLTRERKGGP
GCPNIVLYGVVAGGRDACAALVYSEVRYLPNCGEYVORIMOEKRLVARDNIFL
TRHMSNVGGLCMITLTKETGICPKVLSGPPOLEYLAIRKFFSPGLKIELAVRPH
SAPEYKDEMTVYQVPIHSERCGKQDPSPSPPTSNRLSPKQSSGSGSNGSCPE
DSAGANRRKAMGDEPLVVAFCVRLHNRGNPLVLAKEGLAVGTAAALPILAAVKA
GKSTYEGREIAAELELCTPPDGLVIVVECPDEGLILPCEGNDTRKQAEDADAVA
LVYIAPESVLIDRSRYOQMMEREGPTOLILNENCPVNLHSHKLIOTOLSLIHPI
FQLTSEYKREKESSTLSVPTVRECELIKQLRPKRMQMDPTPCMDCHTILAELEP
SFQESVEETRKVQENPAPAEKRSQPEIVFLGTGSAIIPKIKINVSSTIVNLSPKDS
LDDCGGTGQGLDRLRYGQOIDRYLCSLVAFVSHLADHTGTLNLTLOEHLASLG
LKPQPLIVVAPTDLRAMLOOYHNHCOEILHVSIMIPACIQKGAEVSNITLERLISL

```


Db	2028	ATGGGGAAAGATGCCACCCTCTGATTGATGAAGGCACCTGTGAGAGATGGCTTGGAAAG	2087
Qy	2086	gaagcagctggtgaaaagacacacagcacaacacgtgccaaagccatacgtgtggatgtgcatg	2145
Db	2088	GAAAGCAGTAGAACAACACACAGCACACCTCCAGGCCATTAAATGTGGGATGGCGAGT	2147
Qy	2146	aacgcgagatcattcatgtctgaaccaattcagcagaagcgcctatgtccaaagttccccccttc	2205
Db	2148	AATGGGGGGTTTATCATGCTGTAAGCCACTTCATCTACAGCGGTACGCCAAGATCCCCCTTTTC	2207
Qy	2206	agccccaactcaagcgaagaagctggagctgtgaccttgaccacatgaagtlctgctttga	2265
Db	2208	AGCCCTGACTTCAACAGAGAAAGTTGGCAATCGCCTTTGACACATGAAAGSTGTGCTTGA	2267
Qy	2266	gaacttccaacacatgtcccaagctgatttcccccaactggaagccctgtttgctgcgaatc	2325
Db	2268	GACTTCCCGACAGTGTCCCAAGCTGATTCCTCCACGTGAAGGCCCTGTGTCAGGTGCATT	2327
Qy	2326	gagagagatgaaagacgcagaaggaagacgagacctgcgcagctgcgcgcgcctccctg	2385
Db	2328	GAAAGAGATGTGTGAACGCAGGGGAGCAAGAGAGGACTACGGCTGTGTGCGAGACAGCCCTCTCG	2387
Qy	2386	tccagagagcttgcgcagcgcgctctgtgaagatctggagacctcaacagagaacgycggccacaca	2445
Db	2388	ACC---CAGCAGGCGACAGAGCCACAGAGGACAGAGAAACCCACACAGAAAGGGGGCCACACA	2444
Qy	2446	gagagagccacaca	2456
Db	2445	GATGAACACCA	2455

[illegible]

TITLE	COMMENT
JOURNAL	<p>Macdonald, P., Margulis, N., McEwan, P., McGuirk, A., McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Nachman, A., Nairf, R., Naylor, J., Niloff, M., O'Connor, F., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Rossello, R., Roy, A., Shyam, R., Strange-Thomann, N., Stilwell, J., Stojanovic, N., Stone, C., Strickland, C., Subramanian, A., Torrella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wang, B., Wheeler, J., Wu, Y., Ye, W. J., Zhao, J. and Zody, M.</p>
	<p>Direct Submission Submitted (23-JUL-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Jul 23, 1998 this sequence version replaced gi:3335015. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html</p>
FEATURES	<p>Only the first 118.8 kilobases of this clone are being submitted. The remainder overlaps accession number AC005274 (WIGR project L50).</p>
SOURCE	<p>Location/Qualifiers 1.118786 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="hRPK.597_M.12" /clone_lib="RPCL-11 human BAC library" /map="17" /chromosome="17" 795. .884 /rpt_family="MIR" complement(932. .1000) /rpt_family="MIR" 2044. .2076 /rpt_family="AT_rich" 2237. .2713 /rpt_family="MER34" 2718. .3477 /rpt_family="L2" complement(3478. .3821) /rpt_family="MTL1A" 3822. .4007 /rpt_family="L2" complement(4015. .4281) /rpt_family="MER89" complement(4282. .4321) /rpt_family="(TAA)n" complement(4328. .4608) /rpt_family="AluX" complement(4609. .4895) /rpt_family="MER89" complement(4896. .5186) /rpt_family="AluSp" 5219. .5479 /rpt_family="L2" complement(5481. .5740) /rpt_family="MER34" complement(5963. .6275) /rpt_family="AluSp" complement(6411. .6494) /rpt_family="MER5A" 6830. .7013 /rpt_family="MT1C" 7014. .7318 /rpt_family="AluYb8" 7319. .7581 /rpt_family="MT1C" 7721. .8032 /rpt_family="AluJb" 8039. .8333 /rpt_family="AluSg" 9431. .9460 /rpt_family="AT_rich" 9662. .9829 /rpt_family="MIR"</p>

repeat_region	complement(113032, .11728) /rpt_family="MER39"
repeat_region	complement(11824, .11884) /rpt_family="LTR16C"
repeat_region	complement(11928, .12108) /rpt_family="LTR16C"
repeat_region	12453, .12725 /rpt_family="Aluub"
repeat_region	13377, .13785 /rpt_family="MSRB"
repeat_region	14118, .14153 /rpt_family="CA)n"
repeat_region	complement(15320, .15620) /rpt_family="Aluuo"
repeat_region	15683, .15759 /rpt_family="CA)n"
repeat_region	complement(15838, .15893) /rpt_family="MER1A"
repeat_region	complement(15915, .16080) /rpt_family="MER1A"
repeat_region	complement(16474, .16544) /rpt_family="I2"
repeat_region	17507, .17920 /rpt_family="L1ME"
repeat_region	18271, .18311 /rpt_family="I2"
repeat_region	complement(18356, .18529) /rpt_family="Aluub"
repeat_region	complement(18930, .18560) /rpt_family="CA)n"
repeat_region	complement(18561, .18681) /rpt_family="Aluub"
repeat_region	18760, .19065 /rpt_family="Alusg"
repeat_region	complement(19638, .19896) /rpt_family="MR"
repeat_region	complement(19971, .20260) /rpt_family="L1ME3A"
repeat_region	20263, .20552 /rpt_family="MER58B"
repeat_region	complement(20553, .20816) /rpt_family="L1ME3A"
repeat_region	complement(20821, .21069) /rpt_family="L1M4"
repeat_region	complement(21071, .21100) /rpt_family="TTAA)n"
repeat_region	complement(21101, .21338) /rpt_family="Alusx"
repeat_region	complement(21348, .21598) /rpt_family="L1M4"
repeat_region	21799, .22104 /rpt_family="Aluuo"
repeat_region	22151, .22562 /rpt_family="MS7A"
repeat_region	23978, .24287 /rpt_family="Alusg"
repeat_region	24852, .25199 /rpt_family="I2"
repeat_region	complement(25597, .25985) /rpt_family="MLTID"
repeat_region	25816, .25895 /note="Single-stranded cc complement(25994, .26099) /rpt_family="Aurine-rich"
repeat_region	complement(26122, .26291) /rpt_family="MLTID"
repeat_region	26312, .26337 /rpt_family="CAAAA)n"
repeat_region	26456, .26480 /rpt_family="GGGAA)n"
repeat_region	complement(26576, .26781) /rpt_family="MER6B"
repeat_region	complement(26912, .26990)

/protein_id="AAK9420.1"
 /db_xref="GI:13540342"
 /translation="MVALRSLRLPLGLRTWSOGSARRRPPKDPDLRLHLRTREKRGPG
 GPGPTVYLVQVAAAGRDAGALVYFSEYNRLFCGSGVORLMOEHKLTQVARDNIFL
 TRMWSVGVGLCGMILFKETGLKCKVSGPOLEKFLKAIKJFSEGLKGLAVRPH
 SAPPKDKETMTVYQVPIHSERRCKQOPOSPTSPRLSPKOSDSGSANOCPE
 DSSAGNRKAKGRDPSLVAFVFCVILKRGKGLFLAKKEJLPGVTAIAPIIAVND
 GKSTVYBGRLEIAEELCTPPDPGLVFLVIFSCBPDEGFLPICENDTFERRVQAEADAPVA
 LVNIAIPESVLIDRSYQVMMRFGPDQOHLINNCSPVHNLNHSKLTQTLSDHPD
 LPVLTISFSEKEEGSTLSVPTVGRGELKLTQLRPREMQRTLDNCDEFIAELSLP
 SFQSVSEYERKNVQENAPAPAKESQSYEIVFLGSAIPKIRVSTLVNLSPKSV
 LLDGEGTSGOLCHVQCOIDRYLCSTAFVSHLADHNDTGLNLTQREHALASIG
 KPRPLLVAIPQLRAWLQYHNHCQITLHVSVITPKCKLDKGESNTLBERLISL
 LETDDEEFQTCILVRHCKHAFCALVHSSQMWVYSGDTPCEALVQMGDATALHE
 ALTDDEEAEVEKTHSTSOAILNVGRNMAEFIMLNHFSQRYAKIPLFSPDENKVG
 IAFIMHFCFGDFTVPLKILPPLKALFAGDIEEVEVERERREIRLVRAALLTQADSP
 EDRPOOKRAHNTDEPHSPQSKESVANTLCARV"
 BASE COUNT 8641 a 8030 c 8142 g 9780 t
 ORIGIN
 Query Match 5.8%; Score 144.4; DB 94; Length 34593;
 Best Local Similarity 79.1%; Pred. No. 1.1e-23;
 Matches 201; Conservative 0; Mismatches 41; Indels 12; Gaps 2;
 Oy 2237 cctltgaccacatgaaggctcgtcctltgtgagacattccacaatggcccaagctgattccccc 2296
 Db 30389 CCTGCTGTCCTTCACAGAGCTGCTGTGGAGACTTCCGACAGTGGCCCAAGCTGATTTCCCC 30448
 Oy 2297 cacttgaagccctcttctctcgtcgacatccgagagatgtgaagagccgagggagaaagcggg 2356
 Db 30449 CACTTGAAGCCCTGTTTCCAGTGCATTGGAAGAATGTGTGAAGCCAGGAGGAAGAACAGG 30508
 Oy 2357 agctcgagcaggtgctggggcgcccccctctgtccaggagagctgcaagcgagcctctgagagtg 2416
 Db 30509 AGCTACGGCTGTGCTGGACAGCAGCCCTCTCTGACC---CAGCAGGCGACAGCCACAGAGGACA 30565
 Oy 2417 gggagcctcagcaagaagcgggcccaacaagaaga-----gccacaagcgcaagaag 2467
 Db 30566 GAGAACCCCAACAGAAAGGGGGCCACACAGATGAAACACACAGCCCAACAGAGCAAGAAAGG 30625
 Oy 2468 tcagagcccaagtga 2481
 Db 30626 TCAGAGCCCAAGTGA 30639
 RESULT 11
 LOCUS AF215894 2690 bp mRNA INV 15-AUG-2000
 DEFINITION Drosophila melanogaster juvenile hormone-inducible protein 1
 (Uhl-1) mRNA, complete cds.
 ACCESSION AF215894
 VERSION AF215894.1 GI:9802010
 KEYWORDS
 SOURCE
 ORGANISM
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 2690)
 Dubrovsky,E.B., Dubrovskaya,V.A., Bialderback,A.L. and Berger,E.M.
 The isolation of two juvenile hormone-inducible genes in drosophila
 melanogaster
 Dev. Biol. 224 (2), 486-495 (2000)
 JOURNAL MEDLINE 20387157
 REFERENCE 2 (bases 1 to 2690)
 AUTHORS Dubrovsky,E.B., Dubrovskaya,V.A. and Berger,E.M.
 TITLE Direct Submission
 JOURNAL Submitted (13-DEC-1999) Biology, Dartmouth College, Hanover, NH
 03755, USA
 FEATURES
 source
 1..2690
 Location/Qualifiers
 /organism="Drosophila melanogaster"

Query Match	Best Local Similarity	Score	DB	Length
Matches 475; Conservative 0; Mismatches 436; Indels 30; Gaps 5;	5.4%; 50.5%;	133.4;	5;	2690;
1438	taaccagaatacatcttcctctggaacagggctgccatcccgatgaagattgaaatgctc	1497		
1456	taaccgagatgaattttcttgccgacagcgctccatcccaaaagacggcaaacg	1515		
1498	agtgcaacactgttaaacataagcccgcaacgctctcgtcctgactgctgtagagc	1557		
1516	acctccattcttgatpagaagcctgcaaatgtgctgcttgcgtttttggattgttgagaagaa	1575		
1558	acaattgagcagctgctgcgcgtacatacgaagacag--gtgaaagggctctgggaccc	1614		
1576	acttatggccaattgtatcgacttatntgacacgaaagggcagctgattttggccag	1635		
1615	ctgagctgctgtgttctgtcccaactgacgacagatcacacacgggcttgcaagtac	1674		
1636	ctacagggcagatttattgatcattccatttgcatcgccgacacacacatttgattggtctg	1695		
1675	tgctgcgagaggaacgcgccttgcatcctttgggaaagccgcttcaccccttctgtg	1734		
1696	ctccgggaaag-----gagccaattttaaacttagggcagaccacttatttctg	1743		
1735	gttgccccaacagctcaaacgctggtctcagcagctacacacacagtgccaggagctc	1794		
1744	ctggcctcccgccaatttgaaccttggtt---gagcttctacaaatcgcaaatagaaac	1800		
1795	ctgcaacacatcagatgattctctgccaatgcttcacgaagggcctgagatctccagt	1854		
1801	gtagaagatgcttatctactagctgagcgaacgctgaacctcttgtaagctccgagcggt	1860		
1855	cctgcaagaggaagattgatacgtcgtctgttgcaaacatgatttggaaagattcag	1914		
1861	gaacaagttgaacgcttagaattatgctcaatnccacctgcttagtaggacacgcccc	1920		
1915	acctctgctgtagcgaacgcaacgacgattgctgtgctgctgtagcacacctgagc	1974		
1921	aaacttttcggaatgaacctaaccttgccg-----gcaaacacacatagccgaacc	1971		
1975	tggaagtgatctatccggggacacacatgcccctcgagagctctgtccggaatgggaa	2034		

5;

Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Fitze, E., Galle, R.F., Gary, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Idagwam, C., Jalali, M., Kruse, D., Li, P., Matzel, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, R.A., Nunoo, J., Pacle, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Striplston, M., Strong, R., Swirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 2R, region 47A-47B unpublished

TITLE
JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 175682)
Celnikier, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Chame, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomolan, M.A., Madada, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

Direct Submission
Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 2, 2001 this sequence version replaced gi:5670592.

Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES

source
1. 175682
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="47A-47B"
/clone="BAC48F07 (D625)"
/clone_jib="RPC1-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6)"

BASE COUNT 49836 a 38215 c 38034 g 49597 t
ORIGIN

Query Match 5.4%; Score 133.4; DB 4; Length 175682;
Best Local Similarity 50.5%; Pred. No. 3.6e-21;
Matches 475; Conservative 0; Mismatches 436; Indels 30; Gaps 5;

QY 1438 TACCAGAAATATTTCTTGGGCAAGGCTCTCGATTCCCAATTAAGCGCGAAGCGG 108838
DB 108897 TACCCGAAATATTTCTTGGGCAAGGCTCTCGATTCCCAATTAAGCGCGAAGCGG 108838
QY 1498 AGTGCACACTGTCAACATAAGCCCGACAGCTCTCTACTGACTGTGTGAAGGC 1557
DB 108837 AGCTCATTCTGATCAGAGCTGCAATTAATGCTTGTGATTTGTGAGAGAGA 108778
QY 1558 ACATTGGGCAAGCTGTGCGCTCACTAAGAGAGCAG--GTGAGAGGCTCTGGGAC 1614
DB 108777 ACTTATGCTCAATTTGTAAGTATATGAGACAGAAAGGCGAGTATCTTCCGCG 108718
QY 1615 CCGAGCTGTGTGTGTGTCACACTGTGAGAGAGATCACCACAGGCTGTGCAAGTAC 1674
DB 108717 CTACAGGCGAATTAATCCATTGCTGTCGACACCACTAATGATGATTTGGTCTG 108658
QY 1675 TTGCTGCAAGAGAGAGCGCTGTGCATCTTTGGAGAAAGCCGCTTCACTTGTCTG 1734
DB 108657 CTCGCGGAAG-----GAGGCAATTAACCTAAGGCGAGACCCCACTTATTCG 108610

QY 1735 GTTGCACCACCAAGCTCAAGCTGTGCTCCAGCTACCAACAGTGCAGGAGTGC 1794
DB 108609 CTGGCTCTCCGCAATTAATGAACCTTGTT--GGAATTCTACAAATGCAATAGAAC 108553
QY 1795 CTGCACCACTAGTATGATCTCTCCCAATGCTTCAAGAGGAGTGTAGATCCAGT 1854
DB 108552 GTAGAAAGATGCTATTAATCTTACTGAGCAAGGTAAGTCTTGAAGTCCCTGAGCGT 108493
QY 1855 CCTGAGTGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1914
DB 108492 GAACAAATGTAAGCTTACGATTAATGATTAATGATTAATGATTAATGATTAAT 108433
QY 1915 ACCTCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1974
DB 108432 AACTCTTCTGCAATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 108382
QY 1975 TGAAGATGTCTATCTCCGAGGACCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2034
DB 108381 GTCAAGATCTTACAGCGCGCACTATGATGATGATGATGATGATGATGATGAT 108322
QY 2035 GATGCACCTCTCTAT 2094
DB 108321 GACTCCACGCTTCTTATTCACAGCGCAATGATGATGATGATGATGATGATGAT 108262
QY 2095 GAAAGACACACAGCAAGCTCCCAAGCATCAAGCTGCTGCTGCTGCTGCTGCTGCT 2154
DB 108261 TTAAGACCGCAAGCAAGCTCCGCTTATTCACAGCGCGCAATGATGATGATGAT 108202
QY 2155 TCCATATGCTGAACCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAG 2211
DB 108201 CATGATCTCTCTCCACTTTCGACGCTTACGATTAATGCTGCTGCTGCTGCTGCT 108142
QY 2212 AACCTGAGAGAAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2271
DB 108141 GAGGATATCAACAGATGCTCAATGATGCTTGTGATTAATGATGATGATGATGAT 108082
QY 2272 CCAACATGCTCAAGCTGATCTCCCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAG 2331
DB 108081 CAGCACTACCAAGCTTACCTCCGCTTTCGATGATGATGATGATGATGATGAT 108022
QY 2332 ATGAG 2372
DB 108021 CTGAGACAGCGGCTGTCAAGCGGAGCTGAACAGAGAGAGAGAGAGAGAGAG 107981

RESULT 14
AC007352 190574 bp DNA HTG 24-FEB-2000
LOCUS AC007352
DEFINITION Drosophila melanogaster chromosome 2 clone BACR19J16 (D626) RPC1-98
19.3.16 map 47A-47B strain y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 36 unordered pieces.
ACCESSION AC007352
VERSION AC007352.8 GI:7025687
KEYWORDS HTG; PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 190574)
Celnikier, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Chame, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B., Richards, S., Sethi, H., Swirskas, R.R., Wan, K.H., Webster, D., Woolley, P., Yang, S., Yee, M., Yu, C., and Rubin, G.M.
Sequencing of Drosophila melanogaster
2 (bases 1 to 190574)
TITLE
JOURNAL
REFERENCE
AUTHORS

OY	1915	accitctgcgtgcggcactgcagcagatcgcttggctctgtgcctcgtgtgtgcacccctctgac	1974
Db	133277	AACCTTTTGGAAATAGCCTTAACCTTGGGG-----GCCAAACACAAATAGCCAAACC	133327
OY	1975	tggaaagtggtctatctccggggagacacatctgccttcgagcgtctgtctcggatgggaa	2034
Db	133328	GTCACGATTCACCTTACAGGGGGCAGCACTATGCCATGTCAGAGATGATGATCTGGGGCGCG	133387
OY	2035	gatgcacccctccctgtactcaatgaagccaccctctgaagatgatttggaaaggaagcaatg	2094
Db	133388	GACTCCACCGTGTCTTATTTCACGAGCGCAGCAATGAGGATGACTTTGAGAGGAGGCGCGCG	133447
OY	2095	gaaagacacacagacacacgaagctcccaagccatcagcgttggtggatctcgatgaacgcgag	2154
Db	133448	TTTAAGACGCGCACAGACCGGTGTGCGAGCGTATTTCAGCAGGGCGGGAATATGAACGCCCGCG	133507
OY	2155	ttcattatgcttgaaacactctcagccagcgtatctgccaagtctccctctt--cagcccc	2211
Db	133508	CATACGATTTCCTCCCATTTTTTCGACGGCGTTACCGCTAAGTGCCCGCCGATTAACCGAGTCAAT	133567
OY	2212	aacttcagcgggaagtggtgagtgctctttgacccaatgaagtcctgctcttgaacatt	2271
Db	133568	GAGCATATGCCAAGCAGATTGCAATTAACCTTTGTAATATATGGAAGTAAACCTTGAAGATTCTG	133627
OY	2272	ccaacaatgcccacaagctgattctcccccacttgaagaagcccttcttgcctcgtgcacatcgaag	2331
Db	133628	CAGCACTACCAACAAGCTCTACCCGCCCTTTTCGGGATGTACGCCGAGTACACGGAAAGAA	133687
OY	2332	atggaagagcgcagaggaagaagcgggaagctctgcgcagatgctg	2372
Db	133688	CTGGAGCAGCGGGGCTGTCAAGCGGGAGCTGAAGACGAGACG	133728

RESULT	15
AE003830/c	
LOCUS	
DEFINITION	AE003830 261846 bp DNA INV 04-OCT-2000
ACCESSION	Drosophila melanogaster genomic scaffold 142000013386047 section 13
VERSION	AE003830 AE002787
KEYWORDS	AE003830.2 GI:10727666
SOURCE	HTG.
ORGANISM	fruit fly. Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscocornipha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 261846)
AUTHORS	Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,

REFERENCE
AUTHORS

1 (bases 1 to 261846)

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanalides,P.G., Scherier,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brannon,R.C., Rogers,Y.H., Blazer,G.G., Champe,M., Pfeiffer,B.D.,
Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Milos,G.L., Abill,J.F., Agbayani,A., An,H.J.,
Andrews-Pranckoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borokova,D.,
Botchan,M.R., Bouck,J.R., Brokstein,P., Brothier,P., Butts,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,I.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Foster,C., Gabrielian,A.E., Gang,N.S.,
Gelbart,W.M., Glasser,K., Glodok,A., Gong,F., Gorell,J.J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Iobagvam,C., Jatali,M., Kalush,F., Karpov,G.H., Ke,Z.,
Kernion,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kraivitz,S., Kulp,D., Lai,Z., Lascko,P., Leli,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Matel,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,

Morris J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
 Muzny, D.M., Nelson, D.L., Nelson, D.R., Palazon, K.A., Nixon, K.A.,
 Nusskern, D.R., Paclob, J.M., Palazon, M., Piltman, G.S., Pan, S.,
 Pollard, J., Puri, V., Reese, M.G., Renelt, K., Remington, R.,
 Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Klamos, I.,
 Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
 Stapleton, M., Strong, R., Sun, E., Svrtkas, R., Tector, C., Turner, R.,
 Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A.,
 Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T.,
 Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yen, R.F.,
 Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
 Zhong, F., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
 Gibbs, R.A., Myers, E.W., Rubin, G.M., and Venter, J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006
 2 (bases 1 to 261846)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 On Oct 9, 2000 this sequence version replaced gi:7303755.
 Location/Qualifiers
 1..261846
 source

mRNA
 join(960. ,1069,1135. ,1237,1290,1357,1416. ,1731,2656. ,2759,3801. ,4142,4200. ,4354,4727. ,4835,6101. ,6289,6491. ,6941,6999. ,7841,7921. ,8315)
 /gene="pfk"
 /product="CT13302"
 /db_xref="FLYBASE:FBan0004001"
 /db_xref="FLYBASE:FBgn0003071"
 <960. ,>8315
 /gene="pik"
 /note="CG4001"
 /map="46E3-46E6"
 /db_xref="FLYBASE:FBan0004001"
 /db_xref="FLYBASE:FBgn0003071"
 join(1026. ,1069,1135. ,1237,1290,1357,1416. ,1731,2656. ,2759,3801. ,4142,4200. ,4354,4727. ,4835,6101. ,6289,6491. ,6941,6999. ,7841,7921. ,8059)
 /gene="pik"
 /note="pik gene product [alt 1]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0004001"
 /db_xref="FLYBASE:FBgn0003071"
 /protein_id="AAE58840.1"
 /db_xref="GI:7303792"
 /translation="MHSLKRFVFKLRIPIEINGRIPIDCHPHGPTTFRLEISNKPPIRKQLTFPNIGIOCTSRSHLCCPRDISGNTLTVKKNCKRKHCKLSDSDQKNDRSRGERIKQKDSHQRGKPIINLHNGLFNAVSEKNAAKRSKASRKCGKGVDELRCRLRMQVDYIDVHPKPKDKGLAVFTSGDSQSNAAVRAACVRAIYLGVKYVIREGKYGVGDGDCIOEAMNASVSIIHRGCTIIISARCODFREOGRLKANNLIOGITNNLVYIGDGSGLTGANLPROEWSLDELVTNKLTITTEOOEKFNVLHYIGVLSINDCECTGTDTTADLHRIETAIIDAIISSTASHORTFTWEMGRGCGLTVGLGACADAFPIPMAPKVMDEPRDLCSQLAOERBAGRLNVIYABEAMREBHPITRAEDYKATIDERLKHDPARTLGHVDRGNPSAFEDRIILCRKGAETLMLMEATKUSVPVLSLDNOQAVRPLMECEVERTQAAKAKMAEKRMADAVLKRISERENILETKRLTKLPKPNEDADQGKLEJGLAVMHLIGAPACGMMAAVRSVRVAGYVGIINDVEGLIGNVNELGMSDVSNGOGGAGVIGTRKTLPEGKFEIARLKEFKIIGLLIIGFESYHAAGIADQIRSNVPORCIPIVAVIPSTISNNVPTGERSLGDGGLNETTEICDRIROSAGGTRKRVFLEIMGYCGYLATIAGAGCNDAYIYEKFSITNDLOODQYHMAKMAEGSRRLIYLNKASNSSTDFETRLTSEEGKGLFTRMNLTGHMOOGGSFTFETNMDIIPAKCVLWLAQIKANIDANGVNCSPDTATLLGIVSRQFSPVLDAETNFDQIPKQWMLRLPRLIRLAKHDSAYEEGMYITVEECDDVA"
 /db_xref="FLYBASE:FBgn0003071"
 join(2152. ,2759,3801. ,4142,4200. ,4354,4727. ,4835,6101. ,6289,6499. ,6941,6999. ,7841,7921. ,8315)
 /gene="pik"
 /product="CT13275"
 /db_xref="FLYBASE:FBan0004001"
 /db_xref="FLYBASE:FBgn0003071"

2212 aactcagcgagaaatgggaatttcctttgacccacgaagatcgtcttggaactt 2271
 198535 GAGGATATGCAACAGATTTCCAATACCTTTGATATATGGAAGGACCGTTGAGATCTG 198476
 2272 ccacaatgcaccaagtgtatcccccacttgaagccctgtttgtgcgcatacgagag 2331
 198475 CACACTACCACCAAGCTCTACCCGCCCTTTTCCGATGTACGCCGAGTACAGGAAGAA 198416
 2332 atggaagagcgcaggaagaagcggagagctgcgcagatgcg 2372
 198415 CTGGAGCAGCGGGCTGTCAAGGGGAGCTGAAACAGAGACG 198375

Search completed: October 28, 2001, 19:42:20
Job time: 8694 sec

QY	601	gaagagcctctcagcaggtccatgltccagagcgactcttcagctccgagtgcgatagtaaaaat	660
Db	601	gaagagcctctcagcaggtccatgltccagagcgactcttcagctccgagtgtgaatgaat	660
QY	661	gagccacaccttccacacttggtatgaagcagaagaggggtcaaggagactctccctgtc	720
Db	661	gagccacaccttccacacttggtatgaagcagaagaggggtcaaggagactctccctgtc	720
QY	721	gtaagcttcactctgtaagagcttcacttaaagaagagaaactctctgtgtcacaagcaag	780
Db	721	gtaagcttcactctgtaagagcttcacttaaagaagagaaactctctgtgtcacaagcaag	780
QY	781	gagatgvggctccctcagcttgtagagagcgtgcacgtcccatcatgtcgtgtcaagagac	840
Db	781	gagatgvggctccctcagcttgtagagagcgtgcacgtcccatcatgtcgtgtcaagagac	840
QY	841	gggaaagcatalcatctcttgagaagagagattttgtcgtgaagagctgtgtactcttcca	900
Db	841	gggaaagcatalcatctcttgagaagagagattttgtcgtgaagagctgtgtactcttcca	900
QY	901	gactctgtgtctgtcttctgtgtgtgtgaagatgtccagatgtgaagcttcaataacacac	960
Db	901	gactctgtgtgtcttctgtgtgtgtgaagatgtccagatgtgaagcttcaataacacac	960
QY	961	tgtagaagatgcacacttcttcagaggtaccaaaggaagagcagatgtcccccgtgtgctgtg	1020
Db	961	tgtagaagatgcacacttcttcagaggtaccaaaggaagagcagatgtcccccgtgtgctgtg	1020
QY	1021	gttcaacatggtccccaagatctgtgtctgtgtgtgaagagaggttacccaagatgtgtgaagag	1080
Db	1021	gttcaacatggtccccaagatctgtgtctgtgtgtgaagagaggttacccaagatgtgtgaagag	1080
QY	1081	tttgggcccctgcacacccagcactgtgtcccttaatgtgaagactgtgtccactgaatacaact	1140
Db	1081	tttgggcccctgcacacccagcactgtgtcccttaatgtgaagactgtgtccactgaatacaact	1140
QY	1141	cgtcagcccaagatltcaaacacccagcttcaactctacatccaccggagacatcttcccctgtc	1200
Db	1141	cgtcagcccaagatltcaaacacccagcttcaactctacatccaccggagacatcttcccctgtc	1200
QY	1201	accagtttccgtctgtaaagaggggggtcccaaccttaagtgtgtcccatgttcaaggtgaa	1260
Db	1201	accagtttccgtctgtaaagaggggggtcccaaccttaagtgtgtcccatgttcaaggtgaa	1260
QY	1261	tgccccctcgaagtacccagcccgctccccaagagggagtgtgcagaagagatgccataact	1320
Db	1261	tgccccctcgaagtacccagcccgctccccaagagggagtgtgcagaagagatgccataact	1320
QY	1321	tgcaatccctgagaaatcatcaagtgtgaagcgctgtcagacttcccaactcccaagcagaagcg	1380
Db	1321	tgcaatccctgagaaatcatcaagtgtgaagcgctgtcagacttcccaactcccaagcagaagcg	1380
QY	1381	caggaggtacagagagagtgctgcgcagagagccagccccaagcagaagaaagatcagttac	1440
Db	1381	caggaggtacagagagagtgctgcgcagagagccagccccaagcagaagaaagatcagttac	1440
QY	1441	ccgaataatcaacttccctcttggaacaagaggtctgcacatccgaataaagattcgaatatgt	1500
Db	1441	ccgaataatcaacttccctcttggaacaagaggtctgcacatccgaataaagattcgaatatgt	1500
QY	1501	gcacacactgtgtcaacaataaagcccccagacagcttctctgtactgtgactgtgtgaagagcca	1560
Db	1501	gcacacactgtgtcaacaataaagcccccagacagcttctctgtactgtgactgtgtgaagagcca	1560
QY	1561	tttgtggcagctgtgtgcgttcatcttaacggagacaggtgtgacagaggtccctgtgcacctgtc	1620
Db	1561	tttgtggcagctgtgtgcgttcatcttaacggagacaggtgtgacagaggtccctgtgcacctgtc	1620
QY	1621	gcttgttttgtgtcccaacccgcagcagagatctacacacacagggcttgcacagtatcttctgt	1680
Db	1621	actgtgtgtgtgtcccaacccgcagcagagatctacacacacagggcttgcacagtatcttctgt	1680

QY	1681	cagagagagacgcgccttgcgacatcttttgggaaagccgccttcaaccttgcctggctgttgc	1740
Db	1681	cagagagagacgcgccttgcgacatcttttgggaaagccgccttcaaccttgcctggctgttgc	1740
QY	1741	cccaacccgctcaaaagcccttgctccagcagatgacacaaacccagctgcagagagctctcac	1800
Db	1741	cccaacccgctcaaaagcccttgctccagcagatgacacaaacccagctgcagagagctctcac	1800
QY	1801	cacatcagtatgatctctgcacaaatgccttccaggaagggctgcagatctccagctctgca	1860
Db	1801	cacatcagtatgatctctgcacaaatgccttccaggaagggctgcagatctccagctctgca	1860
QY	1861	gtggaagatgatcatcagttcgcctgttctgcgaacatgctgatttggaaaggtttcagacctgt	1920
Db	1861	gtggaagatgatcatcagttcgcctgttctgcgaacatgctgatttggaaaggtttcagacctgt	1920
QY	1921	ctgggtgcggcgcactgcgaagcatgcgttttggcgtgcctgcctggcgacacctccctgcctgaaa	1980
Db	1921	ctgggtgcggcgcactgcgaagcatgcgttttggcgtgcctgcctggcgacacctccctgcctgaaa	1980
QY	1981	gtggtgctatctccggggagacacatgcctccctgcgagagctcctgcctgcgattggggaaagatgcc	2040
Db	1981	gtggtgctatctccggggagacacatgcctccctgcgagagctcctgcctgcgattggggaaagatgcc	2040
QY	2041	accctctctgatcatatgaagccaccccttggaagaatgcttggaaagagcaatgvgaaaag	2100
Db	2041	accctctctgatcatatgaagccaccccttggaagaatgcttggaaagagcaatgvgaaaag	2100
QY	2101	acacacagacaaagctcccaagccatcaagcgttggggatctgcgatatgaaacgvcggagttacat	2160
Db	2101	acacacagacaaagctcccaagccatcaagcgttggggatctgcgatatgaaacgvcggagttacat	2160
QY	2161	atgcgtgaacccacttcaagccagcgtcatatgcgaagctcccccctcttcaagccccaacttcaag	2220
Db	2161	atgcgtgaacccacttcaagccagcgtcatatgcgaagctcccccctcttcaagccccaacttcaag	2220
QY	2221	gagaaagatgaggagtgtgccttgcacacaatgaagatctgccttggagactttccacaatg	2280
Db	2221	gagaaagatgaggagtgtgccttgcacacaatgaagatctgccttggagactttccacaatg	2280
QY	2281	cccaagcgtgattccccacttgaagaccctgttctctggcgaatccagggagatggaagag	2340
Db	2281	cccaagcgtgattccccacttgaagaccctgttctctggcgaatccagggagatggaagag	2340
QY	2341	cgcaagggagaaagcgggaaatcgccggcagaagtgtgcggcggcccccctctccatgcaggagactgca	2400
Db	2341	cgcaagggagaaagcgggaaatcgccggcagaagtgtgcggcggcccccctctccatgcaggagactgca	2400
QY	2401	ggcggcctcgsgagatgaggagccttcaagcaagaagcgggcccacacagagagccacagagcc	2460
Db	2401	ggcggcctcgsgagatgaggagccttcaagcaagaagcgggcccacacagagagccacagagcc	2460
QY	2461	aagaaagctcagaagcccaag	2478
Db	2461	aagaaagctcagaagcccaag	2478
RESULT 3			
AAC76445			
ID	AAC76445 standard; cDNA; 2546 BP.		
XX			
AC	AAC76445;		
XX			
DT	08-FEB-2001 (first entry)		
XX			
DE	Human ORFX ORF2000 polynucleotide sequence SEQ ID NO: 3999.		
XX			
KW	Human; open reading frame; ORFX; detection; cytosstatic; hepatotropic;		
KW	vulnerary; antliporiatic; antiparkinsonian; nootropic; neuroprotective;		
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;		
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;		
KW	antiviral; antibacterial; antifungal; antirheumatic; antihyroid;		

KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage; nocturnal haemoglobinuria; burn; wound;
KW	thrombosis; contraceptive; ss.
XX	
OS	Homo sapiens.
PN	
PD	WO200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
PR	02-APR-1999; 99US-0127636.
PR	05-APR-1999; 99US-0127728.
XX	
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shimkets RA, Leach M;
XX	
DR	WPI: 2000-602362/57.
DR	P-PSDB: AAB42236.
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X,
PT	useful for treating e.g. cancers, proliferative disorders,
PT	neurodegenerative disorders and cardiovascular disease -
XX	
PS	Claim 5; Page 3179-3180; 5507pp; English.
XX	
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC	sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC	antiprolitic; antiparkinsonian; nootropic; neuroprotective;
CC	osteophtic; anticonvulsant; antihypertic; immunosuppressant;
CC	immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
CC	antidiabetic; hypotensive; dermatologic; immunosuppressive;
CC	antimflammatory; antibacterial; antiviral; antifungal; antineumatic;
CC	antihypertic; and antianemic. The sequences can be used for determining
CC	the presence of or predisposition to, or preventing or treating
CC	pathological conditions associated with an ORFX-associated disorder. The
CC	nucleic acids can be used to express ORFX proteins in gene therapy
CC	vectors. The proteins and nucleic acids may be used to treat cancers,
CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,
CC	hyperextension, hypothyroidism, cholesterol ester storage, systemic lupus
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC	nocturnal haemoglobinuria, antineflammatory disease; to enhance
CC	coagulation; to inhibit thrombosis; and as a contraceptive.
XX	
SQ	Sequence 2546 BP; 652 A; 643 C; 686 G; 564 T; 1 other:
XX	
Query Match	71.8%; Score 1782; DB 21; Length 2546;
Best Local Similarity	99.1%; Pred. No. 0;
Matches 1813; Conservative	0; Mismatches 15; Indels 2; Gaps 2.
OY	654 tgaanaatgagccacaccccttcacacatggtcttagccagagaagaagggtgaagacatc 713
DB	251 tgataatcaagtggttttcttcctcaggtgtcgcagagaagaagggtcaggaactc 310
OY	714 cctgtgcgtagcttcatctgtgaagcttaagcttaagaagaagaactt-ctgtgtgca 772
DB	311 cctgtgcgtagcttcatctgtgaagcttaagcttaagaagaagaagaacttctgtgtgca 370
OY	773 aagcaagaagagatggtgctccagttgtggagacgctgcacatgcctccatcatctgtctg 832

Db	371	aagcaagagagatgaggccctcccaattgaggacagctgcacatcgtcccaatcatgctgcgcg	430
QY	833	tcaagagacgggaaaaagcatcatcatctatgaagaagagagatcttctgcttgaagagctgtga	892
Db	431	tcaagagacgggaaaaagcatcatcatctatgaagaagagagatcttctgcttgaagagctgtga	490
QY	893	ctctctcagatctcgtgtgtcgttcttgtgtgtgtgaagatgtccagatgaagaattcatctc	952
Db	491	ctctctcagatctcgtgtgtcgttcttgtgtgtgtgaagatgtccagatgaagaattcatctc	550
QY	953	aaccocatctgtgagaagatggccaccttaagaggtatcaagaagaaagagacatgtgccccgtgtg	1011
Db	551	aaccocatctgtgagaagatggccaccttcaagaggtatcaagaagaaagagacatgtgccccgtgtg	610
QY	1013	ccttgtgtgttcaacatctgagcccaacatctctgtctgtgtgaacagagttccacagcaatgtga	1072
Db	611	ccttgtgtgttcaacatctgagcccaacatctgtctgtgtgaacagagttccacagcaatgtga	670
QY	1073	tggagaagtttgggtcgtgtgaacccacagcactgtgtcctgaatgaagaactgtgtccatgttc	1133
Db	671	tggagaagtttgggtcgtgtgaacccacagcactgtgtcctgaatgaagaactgtgtccatgttc	730
QY	1133	acaaccttcgagacgcacaagaattcaaaacccagcttcaacctcatccaccccgagcatcttc	1193
Db	731	acaaccttcgagacgcacaagaattcaaaacccagcttcaacctcatccaccccgagcatcttc	790
QY	1193	ccctctcaacacagtttccgcctgttaagaagagagggcccaacccctcaagtgtgcctatgtctc	1253
Db	791	ccctctcaacacagtttccgcctgttaagaagagagggcccaacccctcaagtgtgcctatgtctc	850
QY	1253	agggtgaatgtccctccctcaagtaacagctccgtccacagagagaggtgtgcagaaagtatgccca	1311
Db	851	agggtgaatgtccctccctcaagtaacagctccgtccacagagagaggtgtgcagaaagtatgccca	910
QY	1313	ttatacttgcacaatccctgtagaagaattcatctatgtagaggtgtgcaggtcttcccaattccacg	1372
Db	911	ttatacttgcacaatccctgtagaagaattcatctatgtagaggtgtgcaggtcttcccaattccacg	970
QY	1373	aagagctgcagagaggtatcacagggaggtgcgcagagaaacggccacagccccaagcagaaagaaagaa	1433
Db	971	aagaggtgtcagagaggtatcacagggaggtgtgcgcagagaaacggccacagccccaagcagaaagaaagaa	1033
QY	1433	gtccagtaccgcaagaatcatcttccctgtgaacacaggtctctgcatacccgatgaagaattctgaa	1493
Db	1031	gtccagtaccgcaagaatcatcttccctgtgaacacaggtctctgcatacccgatgaagaattctgaa	1093
QY	1493	atgtcagtgccacaacttgttaacataaagcccgacacacagttctctgtctaaccttgcagctgtgtgtg	1553
Db	1091	atgtcagtgccacaacttgttcaacataaagcccgacacacagttctctgtctaaccttgcagctgtgtgtg	1153
QY	1553	aggagcaacttggggcacagctgtgtgccttattatcagagacacaggtgtgtgacagagttccctgtgca	1613
Db	1151	aggagcaacttggggcacagctgtgtgccttattatcagagacacaggtgtgtgacagagttccctgtgca	1213
QY	1613	ccctgtgcgtctgttcttgtgttccaccctgtgacgcagatcaaccacaaggggtctgtccaagta	1673
Db	1211	ccctgtgcgtctgttcttgtgttccaccctgtgacgcagatcaaccacaaggggtctgtccaagta	1273
QY	1673	tcttgcctgcagaaagaaacgcgtcttggcatcttcttgggaaagccgcgttcaaccttctgtctgtg	1733
Db	1271	tcttgcctgcagaaagaaacgcgtcttggcatcttcttgggaaagccgcgttcaaccttctgtctgtg	1333
QY	1733	tgtgtgtccccaacacagcttcaaaacctgtgcctcagcaggtatcaacaacacaggtgtccagaaag	1793
Db	1331	tgtgtgtccccaacacagctcctcaaaacctgtgcctcagcaggtatcaacaacacaggtgtccagaaag	1393
QY	1793	tctctgcaccacatcatgatattctcctgtccaatgtccttcaagaaaggggtctgtgatatctcca	1853
Db	1391	tctctgcaccacatcatgatattctcctgtccaatgtccttcaagaaaggggtctgtgatatctcca	1453
QY	1853	gtccctgcaggtgtgaaagatttatcaagttcgcgtgttgggaacatgttgatttggaaaggtttc	1913

[illegible]

RESULT	4
AAA60207	
ID	AAA60207 standard; DNA; 26664 BP.
XX	
AC	
AAA60207;	
DT	07-DEC-2000 (first entry)
XX	
DE	Human prostate cancer predisposing gene HPC2 genomic sequence.
XX	
KW	Human; prostate cancer predisposing gene; HPC2; chromosome 17P.
KM	gene therapy; peptide therapy; drug design; ds.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	
FT	Location/Oualifiers 910..26039
FT	/tag= a
FT	/product= "HPC2"
FT	/note= "this sequence contains introns"
FT	/transl_except= (pos:23892..23895,aa:Glu) 910..1154
FT	exon
FT	/tag= b
FT	/number= 1
FT	1736..1786
FT	exon
FT	/tag= c
FT	/number= 2
FT	1925..1995
FT	exon
FT	/tag= d
FT	/number= 3
FT	

FT	exon	3025..3089
FT		/tag= e
FT	exon	/number= 4
FT		4361..4418
FT		/tag= f
FT	exon	/number= 5
FT		5582..5650
FT		/tag= g
FT	exon	/number= 6
FT		7075..7194
FT		/tag= h
FT	exon	/number= 7
FT		8186..8244
FT		/tag= i
FT	exon	/number= 8
FT		12878..12936
FT	exon	/tag= j
FT		/number= 9
FT		13032..13104
FT	exon	/tag= k
FT		/number= 10
FT	exon	13756..13868
FT		/tag= l
FT	exon	/number= 11
FT		15283..15378
FT		/tag= m
FT	exon	/number= 12
FT		16278..16416
FT	exon	/tag= n
FT		/number= 13
FT	exon	16496..16588
FT		/tag= o
FT	exon	/number= 14
FT		18583..18701
FT	exon	/tag= p
FT		/number= 15
FT	exon	20349..20445
FT		/tag= q
FT	exon	/number= 16
FT		22172..22310
FT	exon	/tag= r
FT		/number= 17
FT	exon	22879..22917
FT		/tag= s
FT	exon	/number= 18
FT		23045..23154
FT	exon	/tag= t
FT		/number= 19
FT	exon	23795..23895
FT		/tag= u
FT	exon	/number= 20
FT		23973..24093
FT	exon	/tag= v
FT		/number= 21
FT	exon	24354..24432
FT		/tag= w
FT	exon	/number= 22
FT		25026..25170
FT	exon	/tag= x
FT		/number= 23
FT	exon	25812..26036
FT		/tag= y
FT	exon	/number= 24
FT		26447..26455
FT	polyA-signal	/tag= z
FT		
XX		WO20027864-A1.
XX		18-MAY-2000.
XX		
XX		05-NOV-1999;
XX		99WO-US26055.
XX		06-NOV-1998;
XX		98US-0107468.


```
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154038.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155133.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157153.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3 6%; Score 90.2; DB 21; Length 1617;
Best Local Similarity 53.9%; Pred. No. 3,1e-13;
Matches 185; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 1970 ctygcctgaaagtgtgtctctctccggggaacacatgcccctgcaggtctctgcgcatg 2029
DB 1093 cggggttggaataatgtttctctcggggaacacagcctgtctgagatggttgaagctt 1152
QY 2030 ggaagaatgccaacccctcctgacatgaaagccacccctggaagatgttggaaagaa 2089
DB 1153 caaaggaagcaacagttcttatacatgagcgagcttggagatgcgttggtagaagaag 1212
QY 2090 cagtggaagaagacacagacaaagctcccaagccatcagctgtgggagatgcggaatga 2149
DB 1213 ctgtgtctaataaacatcacactaaagaagcaatcaaatcaagtgtatcttaagaagtg 1272
QY 2150 cggagttcatatgctgaaccacttcagccagcgcgtatgccaaggtccccccttcagcc 2209
DB 1273 tataccgaacgcgtactgacacattttagccagagataccaagaatcccggtgatgaag 1332
QY 2210 ccaacttcagcgagaagaagtggagatgtgccttgcacacatgaaggtctcttggagact 2269
DB 1333 aatcacatagcacatatatgcatgttccttgatgatggtagtataaacaatgycggatt 1392
QY 2270 ttccaacaatgcccagaagctgattccccccactgaagccctgtt 2312
DB 1393 tacatgtgtctccaagaatttaccttacttaaaactctatt 1435

RESULT 8
AAC49965
ID AAC49965 standard; DNA; 2853 BP.
XX
AC AAC49965;
XX
DT 18-OCT-2000 (first entry)
```

[illegible]

```
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156556.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 3.6%; Score 90.2; DB 21; Length 2853;
Best Local Similarity 53.9%; Pred. No. 3.9e-13;
Matches 185; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```
OY 1970 ctgctggaaggtgcttcttcctcgagacacatgctgagagcttctgctcgatgag 2029
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2333 cggggctggaagaaatgcttacttcggaagacaaagccttctcctgagatgtgaagctt 2392
OY 2030 ggaagatgccaacctctctgtatcatggaagcacccttgaagatggttgaagaagaag 2089
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2393 caaagagagcaacgcttcttatactatgagcgagctttagagatgctgtgtgaagaag 2452
OY 2090 caatggaagaaagacacacacacacacacacacacacacacacacacacacacacac 2149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2453 ctgtgtgctaaataacacatgacacactaaagaagcaatcaagttgcatcttcagcagtg 2512
OY 2150 cggagctcatatgtgtagaacacttcagcagcgctatgacgaagttccctcttcagc 2209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2513 tataccgaacgcttactgacacattttagccagagatacccaagaatcccggtgatgag 2572
OY 2210 ccaacttcagcgaagaaatggaagttgcttcttgacacatgaagttcgtcttgagact 2269
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2573 aatcacatatgcaacatacatgcatgtgcttcttgatgatgatatataacaatggcgagt 2632
```

```
OY 2270 ttccacaatgcccgaagctgattccccccactgaagccctgtt 2312
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2633 tactatgcttccaagaatttacttacttacttaacttaacttact 2675
```

RESULT 9
AAC89621
ID AAC89621 standard; DNA; 2517 BP.
XX
AC AAC89621;
XX
DT 08-MAR-2001 (first entry)
XX
DE S. cerevisiae YKR079C gene.
XX
XX Yeast; germination; proliferation; essential gene; antifungal agent;
KW insecticide; herbicide; anti-proliferation drug; cancer; psoriasis;
KW restenosis; YKR081C; YFR003C; YGR277C; YGR278W; YKR071C; YKR079C;
KW YKR083C; ds.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200071161-A1.
PD 30-NOV-2000.
XX
PF 12-MAY-2000; 2000MO-US13017.
XX
PR 21-MAY-1999; 99US-0315794.
PR 02-SEP-1999; 99US-0389341.
XX
PA (ROSE-) ROSETTA INPHARMATICS INC.
XX
PI Roberts CJ;
XX
DR WPI: 2001-025092/03.
DR P-PSDB; AAB49964.
XX
XX Identifying antifungal compounds which target yeast essential genes
PT comprises use of novel Saccharomyces cerevisiae essential genes
PT YFR003C, YGR277C, YGR278W, YKR0701C, YKR079C or YKR083C or YKR081C -
XX
PS Example 5; Fig 29; 127pp; English.
XX
XX The present invention provides methods of identifying antifungal agents
CC using the coding and protein sequences of several yeast genes. These are
CC essential for the germination and proliferation of Saccharomyces
CC cerevisiae, and include YKR081C, YFR003C, YGR277C, YGR278W, YKR071C,
CC YKR079C and YKR083C. The sequences can also be used to identify compounds
CC for use as herbicides, insecticides and anti-proliferation drugs which
CC can be used in the treatment of cancer, psoriasis and restenosis. This is
CC because they can be used to identify plant, insect and human homologues
CC of the yeast genes.
XX
SQ Sequence 2517 BP; 881 A; 417 C; 496 G; 723 T; 0 other:

Query Match 2.4%; Score 58.4; DB 22; Length 2517;
Best Local Similarity 51.2%; Pred. No. 4.9e-05;
Matches 168; Conservative 0; Mismatches 151; Indels 9; Gaps 1;

```
OY 2021 tccgagtgagggaagatgcccctctctgatatcatgaaagcacccttgaagatgttgg 2080
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2126 tcgaatatgcatataatgaatctatcatcagaagctacactagaataatcagctac 2185
OY 2081 aagaagaagcagtgtaaaagacacacagcagacacagctcccaagccttcagcgaggatgc 2140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2186 tggagatgcgctgaagaanaaaacacgcacattcaatgaagaacatcggtgttcgaaca 2245
OY 2141 gatatcaacgcgagatcatatgctgcaacacattcagccagcgcatagccaaagttccccc 2200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2246 aatgaatgctcagagttgattcattacaacacatttccagagatatcccaaatgtccccc 2305
```



```
XX MO9947647-A1.
XX 23-SEP-1999.
XX 12-FEB-1999; 99WO-US03307.
XX 18-MAR-1998; 98US-0040961.
XX 06-AUG-1998; 98US-0130114.
XX (PHAR-) PHARMACOEPIA INC.
XX Dama) BB, Horlick RA, Robbins AK;
XX WPI: 1999-610610/52.
XX P-PSDB; AAY28843.
XX New method for expressing genes from recombinant eukaryotic cells,
XX useful for gene therapy -
XX Claim 24; Fig 2; 86pp; English.
XX The present sequence is a DNA encoding Epstein Barr Virus Nuclear
XX Antigen 1 (EBNA 1), which is obtained from commercially available
XX plasmid PCMVEBNA. EBNA 1 protein is used to stably maintain episomes
XX containing EBV origin of replication (oriP) and a gene encoding
XX protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
XX protein are transfected with these episomes to produce recombinant
XX cell lines expressing multiple genes of interest. This provides a
XX rapid and reliable method of stably expressing multiple genes in
XX transfected cells. The episomes are useful in the transfection of genes
XX encoding receptors, transporter proteins, ion channels, adhesion
XX molecules and transcription factors. The episomes carrying desired genes
XX can also be used to transfect cells in gene therapy, antisense therapy,
XX for gene amplification, cell immortalisation, etc.
XX Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other:
SQ
Query Match 1.9%; Score 48.2; DB 20; Length 1925;
Best Local Similarity 58.9%; Pred. No. 0.017;
Matches 83; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 2327 aggaatgagagagcgagcgaggaagcgagctgcgcaggtgcggcgccctcctgt 2386
DB 444 aggaacagagagagagcgagcgaggaagcgaggaagcgaggaagcgaggaag 503
QY 2387 ccaggaagctgcgcagcgagcgaggaagcgagctgcgcaggaagcgaggaag 2446
DB 504 ggcaggaagcagagagagagcgaggaagcgaggaagcgaggaagcgaggaag 563
QY 2447 aggaacacagagcgagaag 2467
DB 564 agggcgaggaagggcgagag 584
RESULT 14
AAAS0254
ID AAAS0254 standard; DNA: 1926 BP.
XX AAAS0254;
XX 07-NOV-2000 (first entry)
XX Epstein Barr virus nuclear antigen 1 protein (EBNA1) DNA.
XX EBV, nuclear antigen 1; EBNA1; episome; transfection; selection;
XX gene therapy; ds.
XX Epstein-barr virus.
XX WO200047778-A1.
```

```
PD 17-AUG-2000.
XX 11-FEB-2000; 2000WO-US03547.
XX 11-FEB-1999; 99US-0249585.
XX (PHAR-) PHARMACOEPIA INC.
XX Horlick RA, Chelsky D;
XX WPI: 2000-515062/46.
XX P-PSDB; AAY95856.
XX Stably transfecting eukaryotic cells with at least one episome for the
XX production of a desired protein in vitro and for gene therapy -
XX Disclosure; Fig 2; 53pp; English.
XX The present sequence is that of DNA encoding the Epstein-Barr virus
XX (EBV) nuclear antigen protein 1 (EBNA1, see AAY95856). EBNA1 is
XX utilised in a novel method for obtaining a eukaryotic cell that is
XX stably transfected with at least one episome. This method involves
XX transfecting a eukaryotic cell with: (1) a first episome comprising
XX an EBV origin of replication (oriP, see AAAS0253), a gene encoding a
XX first protein whose expression results in cell death and a
XX selectable marker for eukaryotic cells; and (2) a second episome
XX comprising an EBV oriP and a gene encoding a second protein, where
XX expression of the second protein prohibits the occurrence of cell
XX death resulting from expression of the first protein to produce
XX doubly transfected cells which also express an antigen that
XX promotes retention of the episomes by the cells. The doubly
XX transfected cells are maintained under conditions in which the
XX first and second proteins and the selectable marker are expressed,
XX and the selective pressure specified by the marker is maintained.
XX Under these conditions, only cells containing both episomes live.
XX Preferably, EBNA1 is expressed from 1 of the episomes, and the
XX protein of interest from the other episome. Either or both
XX epitopes may further comprise a nucleic acid sequence encoding a
XX protein desired to be expressed in the cell (e.g. a therapeutic
XX protein), a nucleic acid encoding an RNA that is not intended to
XX be translated (e.g. a therapeutic RNA), or a DNA sequence used as
XX a tag for the cells. The method is applicable to cell culture or
XX intact organisms, for gene therapy. It allows the rapid
XX establishment of eukaryotic cells that stably and reliably express
XX a gene of interest, using a novel method of selection, and
XX maintenance of that selection without the need for exogenous
XX selection factors, such as antibiotics.
XX Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other:
SQ
Query Match 1.9%; Score 48.2; DB 21; Length 1926;
Best Local Similarity 58.9%; Pred. No. 0.017;
Matches 83; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 2327 aggaatgagagagcgagcgaggaagcgagctgcgcaggtgcggcgccctcctgt 2386
DB 444 aggaacagagagagagcgagcgaggaagcgaggaagcgaggaagcgaggaag 503
QY 2387 ccaggaagctgcgcagcgagcgaggaagcgagctgcgcaggaagcgaggaag 2446
DB 504 ggcaggaagcagagagagagcgaggaagcgaggaagcgaggaagcgaggaag 563
QY 2447 aggaacacagagcgagaag 2467
DB 564 agggcgaggaagggcgagag 584
RESULT 15
AAAT5454
ID AAAT5454 standard; DNA: 2580 BP.
XX AAAT5454;
XX
```


Query Match	2.4%	Score 58.4;	DB 4;	Length 2517;
Best Local Similarity	51.2%	Pred. No. 4.3e-06;		
Matches 168;	Conservative 0;	Mismatches 151;	Indels 9;	Gaps 1

ESULT 3
S-09-050-863-2
Sequence 2, Application US/09050863

```

1  GENERAL INFORMATION:
2  APPLICANT: Lao, Ying
3  APPLICANT: Hiang, Betty
4  APPLICANT: Payan, Don
5  TITLE OF INVENTION: Mammalian Protein Interaction Cloning
6  TITLE OF INVENTION: System
7  NUMBER OF SEQUENCES: 5
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
10 STREET: 4 Embarcadero Center, Suite 3400
11 CITY: San Francisco
12 STATE: CA
13 COUNTRY: USA
14 ZIP: 94111-4187
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 * COMPUTER: IBM PC compatible
18

```

Query Match	1.9%;	Score 48.2;	DB 3;	Length 2580;
Best Local Similarity	58.9%;	Pred. No. 0.0024;		
Matches 83;	Conservative 0;	Mismatches 58;	Indels 0;	Gaps 0;

RESULT 4
S-09-130-114-1/c
Sequence 1, Application US/09130114

```

? Patent NO. 59/680/7
? GENERAL INFORMATION:
? APPLICANT: Horlick, Robert A.
? APPLICANT: Damaj, Bassam B.
? APPLICANT: Robbins, Alan K.
? TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
? TITLE OF INVENTION: From Multiple Transfected Episomes
? FILE REFERENCE: 0867/1DP030J51
? CURRENT APPLICATION NUMBER: US/09/130,114
? CURRENT FILING DATE: 1998-08-06
? NUMBER OF SEQ ID NOS: 36
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 1
? LENGTH: 5452
? TYPE: DNA
? ORGANISM: VEBNA
? IS-09-130-114-1

```

Query Match	1.9%	Score 48.2	DB 2	Length 5452
Best Local Similarity	56.9%	Pred. No. 0.0036		
Matches	83	Conservative	0	Mismatches 58; Indels 0; Gaps 0
QY 2327	aggagatgagagagcgcaaggagagagctgcgcaggtgcggcgccctcctcgt			2386
Db 1978	AGGAGCAGAGAGAGGGCGAGGAGCAGAGAGAGGGCGCAGGAGGAGAG			1919
QY 2387	ccagggaactctgcacagcggcctcgagagatggggaacctcagaagaacggcccaacag			2446
Db 1918	GGCAGAGCAGAGAGGAGGGCGAGGAGCAGAGAGAGGGCGCAGGAGGAGG			1859


```

|||||
Db 361 GGCTGAGATCTCCAGTCTGCACTGGAAGATGATGATTCGCTGTGCAACATGTGA 420
1899 ttggagaaggttcaaacctgtctgtgaggaactgaagatgaggttggctgtgcct 1958
Db 421 -TTGGAAGAGTTTCAACACCTGTCTGTGGGCACTCAAGCATGCGTTGGCTGCGCT 479
1959 ggtcacacctgtgtctggaagtggtctatctccggggaacacatgacctgagagctct 2018
Db 480 GGTGCACACTCTGCTGTGAAGTGTATTTCCGGGACACCATGCTCCCTGCAAGGCTCT 539
2019 gttccggaatggtggaagaatgcaacctctctgtatataagaaccacctgtgaagtgtt 2078
Db 540 GGTCCGAGTGGGGAAGATGCCACCTCTGATACATGAAGCCACCTGGAAAGATGTTT 599
2079 ggaagaggaagcagctggtgaaagaacacacagcacaagctcccaagcctcagctgggat 2138
Db 600 GGAAGAAGGAGACAGTGTGAAGACACACAGCACACGTCCTCCAGCATCAGCTGGGGAT 659
2139 ggcgaatgaacgcggaatcatatgtctgaaccactcagccagcgtatgccaagttccc 2198
Db 660 GCGGATGAACGCGGAGTTATTTATGCTGAACCACTTCAGC--AGGCTATGCCAAGT-CC 716
Qy 2199 cctctcagcccccaactcagcgaagaagtggagtgctt 2240
Db 717 CCTCTCAGCCCACTTCAAGCGAAGAGTGGAGTGTGCTT 758

RESULT 4
Bg396395 827 bp mRNA EST 12-MAR-2001
LOCUS 602459323f1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4581700 5',
DEFINITION mRNA sequence.
ACCESSION Bg396395
VERSION Bg396395.1 GI:13289941
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 827)
AUTHORS NIH-MGC http://mgi.mgi.nhl.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: gcrabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LILCM1302 row: n column: 05
High quality sequence stop: 779.
Location/Qualifiers
source
1..827
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4581700"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 205 a 217 c 226 g 179 t
ORIGIN

```

```

Query Match 28.3%; Score 702; DB 153; Length 827;
Best Local Similarity 97.8%; Pred. No. 66-173;
Matches 796; Conservative 0; Mismatches 10; Indels 8; Gaps 8;

Qy 686 gccagaagaagggtcaggaactctccctgctgtagcttcatctgtaagcttcaact 745
Db 1 GCCAGAGAAGAGGGGTGAGGTACTTCCCTGCTGCTAGCTTCATCTGAACCTTCACT 60
Qy 746 taaagaaggaactctctgtgtctcaagaagaagaatgggtccctccagttggagag 805
Db 61 TAAAGAGAGAACTTCTGTGCTCAAAAGCAAGAGAGTGGGCTCCAGTTGGAGACAG 120
Qy 806 ctgcactctcccaatcatgtctgctcgaagcaggaaagcaatcaatcatgaagaa 865
Db 121 CTGCCATCGCTCCCATCATTTGCTGCTGTCAAGSACGGAAACATCATCATGAAGAA 180
Qy 866 gagaatttggctgaagagctgtgtactctcccaagctcctgtgtgtcttgggtgg 925
Db 181 GAGAGATTTTGGCTGAAGAGCTGTACTCTCCAGATCCTGGTGTGCTTTGTGTGG 240
Qy 926 tagaatgtccagatgaagcttcaatcaaccatctgtgagaatgccaccttcagaagt 965
Db 241 TAGAATGTCCAGATGAAGCTTCAATTCACCCATCTGTGAATGCCACCTTCAAGAGT 300
Qy 986 accaagaagaagcagatgcccccgctggccttgggtgttcaatgagcccaagctgtgc 1045
Db 301 ACCAAGAAAGGCGAGATGCCCGCTGGCTGTGCTTCAATGCCGCCACAGCATCTGTGC 360
Qy 1046 ttgtgacacagcagtaaccagcagtgatggaaggttggccttgacacccaagcttgg 1105
Db 361 TTGTGACAGACAGAGTACAGCAGTGTGAGAGGTTTGGGCTGTGACCCACGACACTTGG 420
Qy 1106 tctgtgaatgaagactgtgctcagttcaaacctctgcagcccaagatccaaccagc 1165
Db 421 TCTGTGATGAGAACTGTGCTTCAATTCACAACTTGCAGCCCAAGATTCAAACCCAGC 480
Qy 1166 tcaactcatccaccgcgggaatcttccctgtgtcaacagtttcgctgtaagaagga 1225
Db 481 TCAACTCATCCACCGGAGATCTTCCCTGCTCACAGTTTCCCTGTGAAGAGAGG 540
Qy 1226 gccccacctcagctgtgcccac-tgttcaaggtgaaatgctctctcaagtaaccagctcgt 1284
Db 541 GCCCAACCGCTCACTGTGCTCATGTCAGGTTGAGGTTGAAATGCCCTCTCAAGTACAGCTCCGT 600
Qy 1285 cccagaaggaggtgagcagaaggatgccat-ttaacttgaatcctgaggaatcaagt 1343
Db 601 CCCAGAGAGGAGTGCAGAGAGGATGCCATTATTACTTGCATTCGAGGAATTCATAGT 660
Qy 1344 tga-ggcgtctgagcttccc-aacttccagcagagcgtgcaagaagtaaca-ggaagaggtgc 1400
Db 661 TGAAGGCGCTGCAAGCTTCCCAAACTTCCAGACAGCTGAGAGATGACAGGAGAGTGC 720
Qy 1401 gcaagcagggccacgcccagcagaagaagaagtcagtaacc-agaatcatcttccctg 1459
Db 721 GCAGGAACGGCCAG-CCGACAGAGAAAGAAAGATCAGTACCAAGAAATCATCTTCTTG 779
Qy 1460 g-aacgggtctgccatcccgatgaagatcgaa 1492
Db 780 GAAACAGAGGCTGTGCATTCGGATGAAGATCGAAA 813

RESULT 5
LOCUS AU124671 762 bp mRNA EST 23-OCT-2000
DEFINITION AU124671 NT2RM4 Homo sapiens cDNA clone NT2RM4000375 5', mRNA
sequence.
ACCESSION AU124671
VERSION AU124671.1 GI:10949387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

BASE COUNT 230 a 286 c 276 g 168 t
ORIGIN

Note: this is a NIH_MGC Library."

Query Match 29.1%; Score 721.2; DB 169; Length 960;

Best Local Similarity 94.7%; Pred. No. 5.8e-178;

Matches 801; Conservative 0; Mismatches 38; Indels 7; Gaps 5;

```
QY 1362 caactccagcagagcgtgcaggaatagagaagagtcgcagcagcagccagccagc 1421
D 1 CAACATCCAGCAGACCGCTGCAGAGTACAGAGAGTGGCGGAGGAGCGCCAGCCGAC 60
QY 1422 agagaaagaagtcagtcacccaatcatctctcttggaagaagtcgcatccgat 1481
D 61 AGAGAAAGAGCTCACTACCCAGAAATCATCTCTTGGAAGAGGCTGCGCATCCGAT 120
QY 1482 gaagatcgaatcagtcagtcacactgtcaacaagaagcccgacagctctctact 1541
D 121 GAAGATTCGAATATGTCAGTGCACACTTGTCAACATTAAGCCCGACAGCTCTCTACT 180
QY 1542 ggaactgtgtgagggacatttggcagctgtgccttaatacggagagcagtgagcag 1601
D 181 GGACTGTGTGTGGGACAGTGTGGGACAGTGTGGGACAGTGTGGGACAGTGTGGGACAG 240
QY 1602 ggtctctggcagccctgtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1661
D 241 GGTCTGTGGGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
QY 1662 ctgcaagatctctgtcagagaagaagcgccttgatctcttggaagaagcgccttca 1721
D 301 CTGCGCAAGTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
QY 1722 cccttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1781
D 361 CCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
QY 1782 gtgcagaagatctctgtcagacaatcagatcagatctctgtcagaagaagggc 1841
D 421 GTGCCAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
QY 1842 tgaatctcagctcgtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1901
D 481 TGAGATCTCCAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
QY 1902 ggaagatctcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1961
D 541 GGAAGATCTCAGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
QY 1962 gcaacccctgtgagaaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2021
D 601 GCACACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660
QY 2022 ccgagatgggaagaatgtcacccctctgatacatgaagccacccctggagaatgtgtga 2081
D 661 CCGAGTGGGGAACGATGCCA-CTCTCATTCATTAAGC--ACCTGGAAGATGTGACGA 717
QY 2082 agaggaagcagtc-ggaagaagacacacacacagtcacacacacacacacacacacac 2139
D 718 AGAGGAAGACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 777
QY 2140 cgaatgaagcgc-ggaatcattatgtcgaacactcagcagcagcagcagcagcagcagc 2197
D 778 CGAGTGAAGCGCGGACATGACATAGCGGACACAGTCCAGCGGCTAGGCAAAAGGCC 837
QY 2198 ccctct 2203
D 838 CAATAT 843
```

RESULT 3
BE795820 BE795820 761 bp mRNA EST 20-SEP-2000
LOCUS

```
DEFINITION 601590856P1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945085 5',
ACCESSION mRNA sequence.
VERSION BE795820
KEYWORDS BE795820.1 GI:10217018
SOURCE EST.
ORGANISM human.
REFERENCE 1 (bases 1 to 761)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: DCTD/BTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LNCM802 row: p column: 14
High quality sequence stop: 761.
Location/Qualifiers
1..761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3945085"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

Query Match 28.6%; Score 709.6; DB 140; Length 761;

Best Local Similarity 98.3%; Pred. No. 6e-175;

Matches 749; Conservative 0; Mismatches 9; Indels 4; Gaps 3;

```
QY 1479 gatgaagatcgaatcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1538
D 1 GATGAAGATTCGAATATGTCAGTGCACACTTGTCAACATTAAGCCCGACACGCTCTGCT 60
QY 1539 actgagctgtgtgtgagggacacatttggcagctgtgcgtcattacggagacagtgga 1598
D 61 ACTGACAGTGTGTGAGGGACAGTGTGGGACAGTGTGGGACAGTGTGGGACAGTGTGGA 120
QY 1599 cagggtcctgtgagccctgtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1658
D 121 CAGGCTCCTGTGGGACACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
QY 1659 gggcttgcagaatctgtgtcagaagaagcgccttgagatctcttggaagaagcgcct 1718
D 181 GGGCTTGCAGATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
QY 1719 tcaaccttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1778
D 241 TCACCTTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
QY 1779 ccagtgcaagagtgctgtcagcacacatcagatcagtcagcagcagtcagcagcagcagc 1838
D 301 CCAGTGCACGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
QY 1839 ggtgtgatctcagatctcctgtcagtggaagaagatgtatcagtcagtcagtcagtcagtc 1898
```


117: gb_est148:*
118: gb_est149:*
119: gb_est150:*
120: gb_est151:*
121: gb_est152:*
122: gb_est153:*
123: gb_est154:*
124: gb_est155:*
125: gb_est156:*
126: gb_est157:*
127: gb_est158:*
128: gb_est159:*
129: gb_est160:*
130: gb_est161:*
131: gb_est162:*
132: gb_est163:*
133: gb_est164:*
134: gb_est165:*
135: gb_est166:*
136: gb_est167:*
137: gb_est168:*
138: gb_est169:*
139: gb_est170:*
140: gb_est171:*
141: gb_est172:*
142: gb_est173:*
143: gb_est174:*
144: gb_est175:*
145: gb_est176:*
146: gb_est177:*
147: gb_est178:*
148: gb_est179:*
149: gb_est180:*
150: gb_est181:*
151: gb_est182:*
152: gb_est183:*
153: gb_est184:*
154: gb_est185:*
155: gb_est186:*
156: gb_est187:*
157: gb_est188:*
158: gb_est189:*
159: gb_est190:*
160: gb_est191:*
161: gb_est192:*
162: gb_est193:*
163: gb_est194:*
164: gb_est195:*
165: gb_est196:*
166: gb_est197:*
167: gb_est198:*
168: gb_est199:*
169: gb_est200:*
170: gb_est201:*
171: gb_est202:*
172: gb_est203:*
173: gb_est204:*
174: gb_est205:*
175: gb_est206:*
176: gb_est207:*
177: gb_est208:*
178: gb_est209:*
179: gb_est210:*
180: gb_est211:*
181: gb_est212:*
182: gb_est213:*
183: gb_est214:*
184: gb_est215:*
185: gb_est216:*
186: gb_est217:*
187: gb_est218:*
188: gb_est219:*
189: gb_est220:*

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 16:52:01 ; Search time 1834.93 Seconds
(without alignments)
12781.158 Million cell updates/sec

Title: US-09-434-382-1

Perfect score: 2481

Sequence: 1 atgtgggcgccttgcctcgcct.....agaagtcagagccacagtga 2481

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estin6:*
69: em_estin7:*
70: em_estin8:*
71: em_estin9:*
72: em_estin10:*
73: em_estin11:*
74: em_estin12:*
75: em_estin13:*
76: em_estin14:*
77: em_estin15:*
78: em_estin16:*
79: em_estin17:*
80: em_estin18:*
81: em_estin19:*
82: em_estin20:*
83: em_estin21:*
84: em_estin22:*
85: em_estin23:*
86: em_estin24:*
87: em_estin25:*
88: em_estin26:*
89: em_estin27:*
90: em_estin28:*
91: em_estin29:*
92: em_estin30:*
93: em_estin31:*
94: em_estin32:*
95: em_estin33:*
96: em_estin34:*
97: em_estin35:*
98: em_estin36:*
99: em_estin37:*
100: em_estin38:*
101: em_estin39:*
102: em_estin40:*
103: em_estin41:*
104: em_estin42:*
105: em_estin43:*
106: em_estin44:*
107: em_estin45:*
108: em_estin46:*
109: em_estin47:*
110: em_estin48:*
111: em_estin49:*
112: em_estin50:*
113: em_estin51:*
114: em_estin52:*
115: em_estin53:*
116: em_estin54:*

REFERENCE 1 (bases 1 to 762)
 AUTHORS Ota,T., Makamatsu,A., Ozawa,M., Ishii,S., Satou,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and Isogai,T.
 TITLE HRI human cDNA project (Ota,T., Makamatsu,A., Ozawa,M., Ishii,S., Satou,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S., Isogai,T.)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel.: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
 FEATURES
 source Location/Qualifiers
 1..762
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2RM4000375"
 /clone_id="NT2RM4"
 /cell_type="teratocarcinoma"
 /cell_line="NT2"
 /note="Vector: pME18S-FRT3; mRNA from uninduced NT2 neuronal precursor cells"
 BASE COUNT 180 a 212 c 202 g 165 t 3 others
 ORIGIN

Query Match 28.1% Score 698.2; DB 107; Length 762;
 Best Local Similarity 97.9%; Pred. No. 5.8e-172;
 Matches 748; Conservative 0; Mismatches 11; Indels 5; Gaps 4;

QY 852 caccatgaagaagaagaagatcttgctgaagaagctgtgtactctccatgactctgtgtgc 911
 |||||||
 Db 1 CACGATGAGAGAGAGAGATTTGGCTGAGAGAGCTGTACTCTCCAGATCCGCTGTCG 60
 |||||||

QY 912 tgcctttgtgtgttagaagatccagatgaagaagcttcatacccatcttgaaatgac 971
 |||||||
 Db 61 TGCCTTTGTGTGTGATGATGTCCAGATGAAGCTTCATTAACCCATCTGTGAAGATGC 120
 |||||||

QY 972 cacccttcaagatgataccaagaagaagcagatgccccctgtgctgtgtgttccatgac 1031
 |||||||
 Db 121 CACGTTTCAGAGGTAACCAAGAGAGGAGATGCCCGCTGCTGTGTGATTCATATGCC 180
 |||||||

QY 1032 cccagcatgtgtctgtgtgacaagatgaccagcagtgatggaaggtttgggctga 1091
 |||||||
 Db 181 CCCAGCATCTGTGCTGTGGAGACAGCAGGTACACAGATGAGATGAGAGTTGGGCTGA 240
 |||||||

QY 1092 caccagacactgtctccgaatgaaactgtgctccatgtacaaacctgtgcagcacaca 1151
 |||||||
 Db 241 CACCGAGACTGTGCTCGAATGAGAACTGTGCTCACTTACAACTTCGAGCCACAA 300
 |||||||

QY 1152 gattcaaacccagatccaacctatccacccgagacatctccctcgtccacagtttcg 1211
 |||||||
 Db 301 GATTCAAAACCCAGCTCAACCTCATCCACCGAGACATCTTCCCGCTCACCAGTTTCG 360
 |||||||

QY 1212 ctgttaagaagaagggcccaccctcagtggtgccatgttcaagggtgaagtcctcctaa 1271
 |||||||
 Db 361 CTGTAAGAAGAGAGGGCCACCCCTCAGTGTGCCCATGCTTCAAGGGTGAATGCTCTCAA 420
 |||||||

QY 1272 gtaccagtcgtccccaagaagaaggtgacagaagatgcatcttattctcaatccgga 1331
 |||||||
 Db 421 GTACCAAGCTCCGTCACAGAGAGAGGTGGCAAGGAGATCCATTATTTCAATCCCTGA 480
 |||||||

QY 1332 ggaattcatagttgagagcgcgcagctcccaacttccagagagcgcgcaggaataag 1391
 |||||||
 Db 481 GGAATTCATATGTAGGGCGCTGCAGCTTCCCACTTCATCAGACCGGTGCAAGATACAG 540
 |||||||

QY 1392 gagagatgcgacagacagggcccccagccagcagagagaagaagaagtcataccagaatcat 1451
 |||||||
 Db 541 GAGGAGTCCGAGAGAGAGCCAGCCCAAGAGAGAGAGAGAGAGTCACTGCCAGGAATCAT 600
 |||||||

QY 1452 ctctcttgaacagaggtctgcatcccatgaagaatgcgaatgtcagtcgacactgtt 1511
 |||||||
 Db 601 CTTCCTTGGAAACAGGCTCTGCATCCGATGAAATTCGAATGTCACTGGC-CACCTGT 659
 |||||||

QY 1512 caacataagcccccagacagctctctgtac-tgagactgtgtgagggcaactttggcagc 1570
 |||||||
 Db 660 CAACATAAGCCCCCGACACAGCTCTGCTACTTGAGACTGTGTGAGAGGACACATTGG--CAC 717
 |||||||

QY 1571 tgtgcgcgtcatgaagacagcaggtgagcaggtgctctggacc 1614
 |||||||
 Db 718 TTNGCCCGCATTTCCAGAGCAAGTGG-CAGGGTCTTGCGAC 760
 |||||||

RESULT 6
 LOCUS BG386090 920 bp mRNA EST 12-MAR-2001
 DEFINITION 602455264F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583679 5',
 mRNA sequence.
 ACCESSION BG386090
 VERSION BG386090.1 GI:13279536
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs.rem@nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LML at:
 http://image.llnl.gov
 Plate: LCM1307 row: p column: 16
 High quality sequence stop: 736.
 Location/Qualifiers
 1..920
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4583679"
 /clone_id="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pOTB1; Site: 1; XhoI; Site: 2; EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAGAG(6). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 196 a 261 c 272 g 190 t 1 others
 ORIGIN

Query Match 27.8% Score 689.4; DB 153; Length 920;
 Best Local Similarity 91.9%; Pred. No. 1.2e-169;
 Matches 786; Conservative 0; Mismatches 56; Indels 13; Gaps 5;

QY 896 ctccagatcctgtgtgtcttctgtgtgtgatagtctcagatgaagaatctcattcaac 955
 |||||||
 Db 2 CTCGAGATCCGTGCTCTGC-TTGTGTGTGTGAGAAATGTCCAGATGAAGCTTCATTCAAC 60
 |||||||

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS
956	ccatctgtgagatgacacacctttagagatgtccaaagaaagcagatgccccgtgacct	1015					
61	ccatctgtgagatgacacacctttagagatgtccaaagaaagcagatgccccgtgacct	120					
1016	tgtgtgttcaacatggtcccccagatctgtccttgttgaacagaggataccagcagtgtg	1075					
121	tgtgtgttcaacatggtcccccagatctgtccttgttgaacagaggataccagcagtgtg	180					
1076	agaagtttgggacctgaaccccagcacttgtctcctgtaatgtgaactgtgtcctcagttaca	1135					
181	acaggttttgggacctgaaccccagcacttgtctcctgtaatgtgaactgtgtcctcagttaca	240					
1136	acccttcgacgcacacaagaattcaaacccagctcaacctatccaccggacatctctccccc	1195					
241	acccttcgacgcacacaagaattcaaacccagctcaacctatccaccggacatctctccccc	300					
1196	tgtctaccagatcttcgcctgttgaagagagcccccactagtgtgtccaatgtgtcgaag	1255					
301	tgtctaccagatcttcgcctgttgaagagagcccccactagtgtgtccaatgtgtcgaag	360					
1256	gtgaatgtccctcctcaagtaaccagctccgtcccaagaagagagtgtgcagaaggaatgcatta	1315					
361	gtgaatgtccctcctcaagtaaccagctccgtcccaagaagagagtgtgcagaaggaatgcatta	420					
1316	ttacttgcgaatcctgaaggaatcatagttgagtcgctgcagcttcccaacttccagacaga	1375					
421	ttacttgcgaatcctgaaggaatcatagttgagtcgctgcagcttcccaacttccagacaga	480					
1376	gcgtgcagaagagtaacagagagagtggtgcagagacggtcccccagcagagaagaagaagtc	1435					
481	gcgtgcagaagagtaacagagagagtggtgcagagacggtcccccagcagagaagaagaagtc	540					
1436	agtaaccaggaatactctccttgtgaagaagagctgtccatccgatgaagaatctcgaatg	1495					
541	agtaaccaggaatactctccttgtgaagaagagctgtccatccgatgaagaatctcgaatg	599					
1486	tcaagtgcacacacttgtcaacataagccccgcagacagctctgtctacttgaactgtgtgag	1555					
600	tcaagtgcacacacttgtcaacataagccccgcagacagctctgtctacttgaactgtgtgag	659					
1556	gcaacttgggagcagctgtgcgcgtacatcaggaagacagagtgaagaagctctgtggcaacc	1615					
660	gcaacttgggagcagctgtgcgcgtacatcaggaagacagagtgaagaagctctgtggcaacc	717					
1616	tgtgtgtgtgtgttgttgttgcaccacgtgcagacagatcaccaacagtgcttgcagaatct	1675					
718	tgtgtgtgtgtgttgttgttgcaccacgtgcagacagatcaccaacagtgcttgcagaatct	768					
1676	tgtctcaagaagaagaagcgccttggatctcttgggaagaacgccttcaaccttgcgtgtg	1735					
769	tgtctcaagaagaagaagcgccttggatctcttgggaagaacgccttcaaccttgcgtgtg	828					
1736	tgtcccccaaccagc 1750						
829	tgcggggccacccggc 843						
7	7						
A0126037	A0126037	712 bp	mRNA	EST	23-OCT-2000		
LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS
A0126037	NT2RM4 Homo sapiens cDNA clone NT2RM4.002610 5', mRNA						
A0126037	sequence.						
A0126037	GI:10950753						
EST.	human.						
ORGANISM	Homo sapiens						
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
1 (bases 1 to 712)							
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,							
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and							
Isogai,T.							

[illegible]

/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGG(5'). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 227 a 259 c 316 g 182 t
ORIGIN

Query Match 27.3%; Score 676.6; DB 139; Length 984;
Best Local Similarity 98.2%; Pred. No. 2.9e-166;
Matches 726; Conservative 0; Mismatches 9; Indels 4; Gaps 4;

```
QY 682 gttagccagagaagggggtcaggaccttccttgctgtaagcttcattcgtgaagctt 741
    |||||||
DB 12 GTTAGCCAGAGAGAGGGGCTCAGGAGCTCTCCCTGCTCCTAGCTTCATCTGAAGCTT 71
    |||||||

QY 742 cacttaagaagagaactcttggtgctcaagaagaagatgggctcccaagttggg 801
    |||||||
DB 72 CACTTAAGAGAGAGAACTTGTGCTCAAGCAAGAGAGAGGCTCCCAAGTTGGG 131
    |||||||

QY 802 acagctgcacatcgctcccatcattgctgctgctcaagaagcggaagaacatcactatgaa 861
    |||||||
DB 132 ACAGCTGCGCATTCCTCCATCATTTGCTGCTCAAGAGCGGGAAGAACATCATTCAATGAA 191
    |||||||

QY 862 ggaagagagaatttgctgaagagctggtactcctcagatccctggtgctgtt 921
    |||||||
DB 192 GGAAGAGAGA-TTTGGCTGAAGAGCTGTGTACTCTCCATCATCTGCTGCTGC-TTTGTG 249
    |||||||

QY 922 gtggtgaagtgcacatgaagaacttcataaaccatcgtgtagaagtgccacttcag 981
    |||||||
DB 250 GTGTGAAGAGTCCAGTGAAGCTTCATCAACCACTGTGTGAAGATGCGACCTTTTCAG 309
    |||||||

QY 982 aggtaccagaagaaagcagatgccccgtgacctgtgtgttccatgagcccaagcattc 1041
    |||||||
DB 310 AGGTACCAAGAAAGGAGAGATGCCCGCTGGCTTGGTGTTCACATGAGGCCACGACATCT 369
    |||||||

QY 1042 gtgctgtgtgacagcaggtaccagcagtgatggaagatttggtgacctgaacccaagc 1101
    |||||||
DB 370 GTGCTTGTGACAGAGGTACCAAGTGTGATGAGAGAG-TTTGGCTGCTGACACCCACACAC 428
    |||||||

QY 1102 ttggtctgaatgaagaactgtgctcagttcacaacctgscagcagaagaattcaaac 1161
    |||||||
DB 429 TTGGTCTGAATGAGAACTTGCTCTCAGTTCACAACTTGCGACACCAAGATTCAAAACC 488
    |||||||

QY 1162 cagctcaacctcaccacccgagacatctccctgtgtcaccagtttcgctgtaagaag 1221
    |||||||
DB 489 CACCTCAACCTCAACACCCGGAGATCTCCCTGCTCAACGATTCCGCTGTGAAGAAG 548
    |||||||

QY 1222 gaagggcccaacctcagttgtcccatggttcagggtgaatgctcctcctaagtaaccagctc 1281
    |||||||
DB 549 GAGGGCCCTCCCTCAGTGTGCCATGTTCAGAGGTGAAGGGTCCCAAGTACCACTC 608
    |||||||

QY 1282 cctcccaagaggaagtgagagagatgcatattactgcaactccttgagatattata 1341
    |||||||
DB 609 CGTCCCAAGGAGGAGTGGCAGAGGAGATGCCATTATTAATTGCAATCTTGAGGAATTCATA 668
    |||||||

QY 1342 gttagagcgtctgagcttcccaactccagcagagcgtgtaggaagtaagaagatgagc 1401
    |||||||
DB 669 GTTAGAGGCGGTG-AGCTTCCCACTTCACAGACAGCTGCAGAGATACAGAGAGAGTGGC 727
    |||||||

QY 1402 caggagcggccagcccccag 1420
    |||||
DB 728 CAGAGGCGCAGCCACAGAG 746
    |||||
```

RESULT 10
BE260495 676 bp mRNA EST 26-OCT-2000
LOCUS BE260495 676 bp mRNA EST 26-OCT-2000

DEFINITION 601150702F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
mRNA sequence.
ACCESSION BE260495
VERSION BE260495.1 GI:9131807
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 676)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
plate: LNCM176 row: d column: 01
High quality sequence stop: 672.

FEATURES
source
1..676
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3503184"
/clone_lib="NIH-MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGG(5'). Library constructed by Ling Hong
in the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH-MGC library."
BASE COUNT 154 a 207 c 176 g 139 t
ORIGIN

Query Match 27.2%; Score 676; DB 165; Length 676;
Best Local Similarity 100.0%; Pred. No. 3.7e-166;
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1093 acccagcaacttgctcgtgaatgagaactgtgctcagttcacaacctgcagccacaag 1152
    |||||||
DB 1 ACCCAGCACTTGTGCTCGAATGAGAACTGTGCTCAGTTCACAACCTTCCGACCCACAAG 60
    |||||||

QY 1153 attcaaacagtgtaacctcaccatccacccggagacattccctgtgcacagttccgc 1212
    |||||||
DB 61 ATTCAAAACCAGTCAACCTCATTCACCCGACATCTTCCCTGCTCACACATTTCCGC 120
    |||||||

QY 1213 tctgaagaagagggcccccaccctcagtggtgccatggttcagggtgaatgcctcctcaag 1272
    |||||||
DB 121 TGTAAAGAGAGAGGCCCCACCTCAGTGTGCCATGTTCAGGTTGAGGTGAATCCCTCAAG 180
    |||||||

QY 1273 taccagctccgtccacaggaagagtgagagagatgcatattactgcaactcctgag 1332
    |||||||
DB 181 TACCAGCTCGTCCCGAGAGAGAGTGGCAGAGGAGATGCCATTATTAATTGCAATCTCGAG 240
    |||||||

QY 1333 gaattcatagttgagcgctgcagcttcccaacttcagagagagcggtgaggggtacagg 1392
    |||||||
DB 241 GAATTCATAGTTGAGGCGCTGCAGCTTCCCACTTCACAGAGAGCGGTGCGAGGTACAGG 300
    |||||||

QY 1393 aggaagtcgagagagagccagccagccagcagagagaaagaagtaacagtaaccagaatc 1452
    |||||||
DB 301 AGCAGTCCGACAGAGCGGCCACCGCCACGACAGAGAAAAGAGTACGATCCAGAAATCATC 360
    |||||||

QY 1453 ttctctggaacagaggtctgcatcccgatgaaagattcgaattgcaagttgcaactgttc 1512
    |||||||
```


CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLM9561 row: c column: 03
 High quality sequence stop: 686.

FEATURES
 source
 1. 812
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3847226"
 /clone_1ib="NIH_MGC_65"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPO6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."

BASE COUNT 200 a 216 c 220 g 176 t

ORIGIN

Query Match 26.7%; Score 662.6; DB 141; Length 812;
 Best Local Similarity 97.4%; Pred. No. 1.3e-162;
 Matches 716; Conservative 0; Mismatches 14; Indels 5; Gaps 4;

QY 724 gccctcatctgttaagcttcaacttaagaagagaacttctgtgtctcaagaagag 783
 |||||||
 Db 1 gccttcacgtgtagcttcaacttaagaagagaacttctgtgtctcaagaagag 60
 |||||||
 QY 784 atgagctcccaagcttggaagagctgcacatgcctccatctgtctgtcgaagagcgg 843
 |||||||
 Db 61 atggccctcccaagcttggaagagctgcacatgcctccatctgtctgtcgaagagcgg 120
 |||||||
 QY 844 aaaaagcactcaatgaagaagaagatlltgcctgaagaagctgtgtactcctccagat 903
 |||||||
 Db 121 AAAAGCATCATCATGATGAGAGAGATTTTGGCTGAAGAGCTGTACTCCTCCAGAT 180
 |||||||
 QY 904 cctgtgtgtcttctgtgtgtagaatgtccagatgaagaagcttcaaccatctgt 963
 |||||||
 Db 181 cctgtgtgtcttctgtgtgtagaatgtccagatgaagaagcttcaaccatctgt 239
 |||||||
 QY 964 gagaatgcaacctttagagagttacaaagaagaagagagagagagagagagagag 1023
 |||||||
 Db 240 GAGAAATGCCACCTTTCAGAGGTACCAAGGAAGAGAGATGCCCGGCTTGTGTG 299
 |||||||
 QY 1024 caaatgcccacacatctgtctgtgtgaagcaggttacccagagtgatgagagagttt 1083
 |||||||
 Db 300 CACATGGCCCGACATCTGTGCTTGTGAGACAGATACAGAGAGTGGATGAGAGGTTT 359
 |||||||
 QY 1084 gggcgtgaacacacagcacttgcctcgtgaatgagaactgtgctcagttcaaaccttcgc 1143
 |||||||
 Db 360 GGGCCTGACACCCAGCACTTGTGCTGAATGAGAACTGTGCTCACTTCACTTCCG 419
 |||||||
 QY 1144 agcccaaatgaatgaacccagcctcaacccatccaccggagacatcttccccctgcacac 1203
 |||||||
 Db 420 AGCCAAATGAATTAACCAAGCTCAACCTCATCCACCCGGACATCTTCCCTGCTACCC 479
 |||||||
 QY 1204 agttccgctgtgaagaagaagagagagagagagagagagagagagagagagagag 1263
 |||||||
 Db 480 AGTTTCGGCTGTGAAG 539
 |||||||
 QY 1264 cctcctcaagtaaccagctcgttcccaagagag - agtgcagagagagagagagagagagag 1320
 |||||||
 Db 540 CTCTCTCAATACACAGCTCCGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
 |||||||
 QY 1321 tgcacatccgag 1380
 |||||||
 Db 600 TCCCAATCTGAGAGAAATTCATAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658
 |||||||
 QY 1381 caagagtag 1440
 |||||||

Db 659 CAGAGATACAG 718
 |||||||
 QY 1441 ccagaatcatcttc 1455
 |||||||
 Db 719 CCAGAAATCATCTCC 733
 |||||||

RESULT 13
 BP794577 953 bp mRNA EST 12-JAN-2001
 LOCUS 602255843F1 NIH_MGC_85 Homo sapiens CDNA clone IMAGE:4339187 5',
 DEFINITION mRNA sequence.
 ACCESSION BP794577
 VERSION BP794577.1 GI:12099631
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC <http://mgi.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: Louis Straudt, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLM9949 row: e column: 12
 High quality sequence stop: 704.

FEATURES

source
 1. 953
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4339187"
 /clone_1ib="NIH_MGC_85"
 /tissue_type="lymphoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lymph; Vector: pCMV-SPO6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.867 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 214 a 255 c 271 g 213 t

ORIGIN

Query Match 26.5%; Score 656.6; DB 169; Length 953;
 Best Local Similarity 97.7%; Pred. No. 5e-161;
 Matches 687; Conservative 0; Mismatches 14; Indels 2; Gaps 2;

QY 594 gattccagaagaagcctctcagcagagctcagtcagagcagatcttcaagactccgaagtcgaa 653
 |||||||
 Db 7 GTGGGAGAGAAAGGCTCTCTCAGAGGCTCACTCCAGAGAGAGAGAGAGAGAGAGAGAG 66
 |||||||
 QY 654 tgaatgaagcaccacacttccacatggtgtgaagcagaagaagaagagagagagagagag 713
 |||||||
 Db 67 TGAATAATGAGCCACACCTTCCACATGATGTGAGCCAGAGAGAGAGAGAGAGAGAGAG 126
 |||||||
 QY 714 cctgtgtgtgaagcttcatctgtgaagcttcaacttaagaagaagaagaacttctgtgtcga 773
 |||||||
 Db 127 CCGTGTGTGAGAGCTTTCATGTGTGAGCTTCACTTAAGAGAGAGAGAGAGAGAGAGAG 186
 |||||||
 QY 774 agcaagaagagatggcctccagcttggaagagctgcacatgcctccatcatctgtctgt 833
 |||||||
 Db 187 AGCAAAAGGAGATGGGCTCCCACTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246
 |||||||
 QY 834 caagagcgggaaagacatcatgaagaagaagaagatttggcgaagagctgtgtac 893
 |||||||

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
COMMENT Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LINC841 row: j column: 23
High quality sequence stop: 732.

FEATURES

source

```
1. /735
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3959926"
/clone_lib="NIH_MGC_21"
/issue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: Placenta; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 168 a 206 c 203 g 157 t 1 others

ORIGIN

Query Match

Best Local Similarity 26.4%; Score 655.6; DB 141; Length 735;

Matches 719; Conservative 0; Mismatches 14; Indels 5; Gaps 5;

```
QY 1190 tcccctgtaccagttccgctgtgaagaaggagggccaccctcaagtgtcccatgg 1249
DB 1 TCCCCCTGCTCACCACTTCCCGCTGTAAGAAAGAGGGGCCACCTCAGTGTCCCATGG 60
QY 1250 ttcaagggtaatccctcccaagtaccagctcgtcccaaggaggaggtgacaggatg 1309
DB 61 TTCAGGATGATGCTCTCCTCAAGTACCAAGCTCCGTCACAGAGGAGTGGCAGAGGGATG 120
QY 1310 ccatltaactt-gcaatccctgaggaatcatagttgagcgctgcagcttcccaattc 1368
DB 121 CCATTATTACTTNGCAATCCTGAGGAATTCATAGTTAGGCGCTGCACGCTTCCCAACTTC 180
QY 1369 cagcagagcgctgcagaggtacaggaagagtgcgcaagacggccagcccaagagaaa 1428
DB 181 CAGCAGAGCGCTGCAGAGGTACAGAGAGAGTGGCAGAGCGCCAGCCACAGCAGAGAAA 240
QY 1429 aagaagcagttaccagaatacatctcccttggaacaggggtctccatccgatgaagatt 1488
DB 241 AGAAGTCAGTACCCAGAAATCATCTTCTTGGAACAGGGGTCTGCATCCCGATGAAGATT 300
QY 1489 cgaatgtcagttgcacactgtlcaacataagcccgacacagctctctgctactggaactg 1548
DB 301 CGAAATGTCAAGTGCACACTTGTCAACATTAAGCCCGACACGTCCTGTGACTGACTGT 360
QY 1549 ggtgagggcacatttggcagcgtgtgcgtcatltaacgagaaccagtgagaggtctctg 1608
DB 361 GGTGAGGGCACATTTTGGCAGCTGTGCCGTCATTAAGGAGACAGGAGGTCTCTG 420
QY 1609 ggcacccctggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1668
DB 421 GGACCCCTGGCTGCTGTG-TTGTTCCACCTGCACGACAGATCACCAACAGGGCTTGCCA 479
QY 1669 agtatcttgcagagagaacgcgccttggcatcttgggaaagcgcttcaaccttgg 1728
DB 480 AGTATCTTGTGTCAGAGAGAACCGGCTTGCAATC-TTGGGAAAGCGGCTTCAACCTTTG 538
QY 1729 ctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1788
```

```
DB 539 CTGGTGGTGGCCCCCAACAGCTCAAGCCTGGCTCCAGCAGTACCAACACAGTGCAG 598
QY 1789 gaggtcccgaccacatcaatgatattcctgcacaatgcttcaaggagggtgaatc 1848
DB 599 GAGGTCTGACACACATCATGATGATTCCTG-CAATGCTTACAGGAAGGGCTGAGATC 657
QY 1849 tcaagtcctgcaatggaagaattgatcagtcgtgtgtgcaacatgtattt-ggaaga 1907
DB 658 TCCAGTCTTCAGTGGAAAGATTGATCAGTTGCTGTGGGAACATGTGATTTGGGAGA 717
QY 1908 gttcagacctgtctggt 1925
DB 718 GTTTCAGAACTGTCTGTGT 735
```

Search completed: October 28, 2001, 18:35:36
Job time: 6215 sec

Mon Oct 29 07:30:10 2001

us-09-434-382-1.rst

Page 15

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 18:35:41 ; Search time 3261.8 Seconds
(without alignments)
11765.125 Million cell updates/sec

Title: US-09-434-382-1

Perfect score: 2481
Sequence: 1 atgttgagcgttgcgcgt.....agaagtcagagccagatga 2481

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 8

Total number of hits satisfying chosen parameters: 2427854

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rtd: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rtd: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

Seq ID 1
Oligo search
at first
8 news

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2481	100.0	2908	89	AF304370 Homo sapi
2	2379	95.9	2997	91	BC001939 Homo sapi
3	2379	95.9	3006	91	BC004158 Homo sapi
4	2022	81.5	2976	89	AK001392 Homo sapi
5	1667	67.2	2908	89	AF308698 Pan trogl
6	1266	51.0	2893	89	AF308694 Gorilla g
7	247	10.0	118788	85	AC005277 Homo sapi
8	142	5.7	740	89	AF304371S2 Homo sapi

OY	481	atagaactgtgctgtgvgcccacctctgccccaagatcagagatgaaacatgacagt	540
Db	481	ATAAACATGGCTGTGGCGGCCCCACTCGTGGCCAGATACGAGAGTGAACCATGACACTT	540
OY	541	taccagttcccataacagtgacagagagagaggaagcaccaaccatggcagagtcca	600
Db	541	TACCAGATCCCATACACACTGTAAACAGAGAGAGGGGAAACACCAACATGGCAGAGTCA	600
OY	601	gaaagcctctcagcagagctcagtcacagagcagctctcagctccagtcgaatgaaat	660
Db	601	GAAGAGGCTCTCAGCAGGCTCACTCCAGAGGCATCTTCAGACTCCGAGTCCGATATGAAT	660
OY	661	gagccacactccacaatggtgtgtaagccagagaagaggttcagagagactcttcctgtgc	720
Db	661	GAGCCACACCTCCACATGAGTGTGTACCGAGAGAAGGGGCTCAGGAGCTTTCCTGTGTC	720
OY	721	gtagcttcacatctgttaagctcaactaaagagagaaactcttvtgtctaaagcaag	780
Db	721	GTACTTTCATCTGTAAAGCTTCACTTAAAGAGGAAACTTCTTGCTCTCAAAAGCAAG	780
OY	781	gagaatggccctcccaagtlttgagaaagctgcacatgcctcccatatgtcgtcgttaagac	840
Db	781	GAGATGGCCCTCCCAAGTTGGAGACGTGCACTGCTCCCATATGTGCTGTCAAGAC	840
OY	841	gggaaagacatcaactcagtgaagaaagagagatttgcgtgaagagctgttactctcca	900
Db	841	GGGAAAGCATCACTCATGTGAAGAGAGAGATTTTGGCTGAAGAGCTGTACTCTCTCA	900
OY	901	gattccgtgctgctcttltgtgtgtagaatgtccagatgaaagcttcaatcaaccatc	960
Db	901	GATCCTGTGCTGCTGCTTTGTGGTGTAGAGATGCCAGATGAAGCTTCATTCAACCCATC	960
OY	961	tgttaagaatccacactttaagagatccaaaggaagagagatgccccgtgtccttggt	1020
Db	961	TGTAGAGATGCCACCTTTCAGAGGTACCAAGGAAGAGGAGATGCCCTGTGCTGTG	1020
OY	1021	gttacaatggccccaagacatctgtctgtgtggaacagagaatacagcagtgatggaag	1080
Db	1021	GTTCAACTGGCCCCAGCATGTGTGCTGTGTGGACAGCAGGTACCACTGATGTGAGAGG	1080
OY	1081	tttggcctgacacccaagcacttgttctgtgaatgagaaactgtgctcagttcaaacctt	1140
Db	1081	TTTGGCGCTACACCCAGCACTGTTGCTCGAATGAGAACTGTGCTCAGTTACAACTT	1140
OY	1141	cgagccacaagaattcaaacccaagctcaacatctccacccagagatcttccctgtctc	1200
Db	1141	CGCAGCCACAAGATTCAAACCCAGCTCAACTATCTCCACCCGAGCATTTCCCTGTCTC	1200
OY	1201	accagttctcgctgttaagagagagggcccaaccctaaagtgtcccatgtgtcaagtgtaa	1260
Db	1201	ACCAAGTTCCGCTGTAAAGAGAGGGCCCCACCTCATGTGTCCCATGGTTCAAGGTTAA	1260
OY	1261	tgccctctcaagtiacagcgtctgccacagagagagtggaagagatgacatattact	1320
Db	1261	TGCCCTCTCAAGTACACAGCTCCGCTCCCAAGAGAGAGTGCAGAGAGGATGCCATTATTACT	1320
OY	1321	tgcgaatccctgaagaaatcagaatgttgagggcgtgcagacttcccaactccagacagaggt	1380
Db	1321	TGCATCTCTAGGAATTCATATGTTGAGGGCGTGCACACTTCCCAACTCCACAGAGCGTG	1380
OY	1381	cagagatcacagagagagtgtagcagagacggcccaagccccaagagagaaaaagatcaatrac	1440
Db	1381	CAGAGATCACAGAGAGAGTGGCGCAGAGCGGCCCAAGCCCAAGCAGAAAAAGAGTCAAGTAC	1440
OY	1441	ccagaatatcatcttcccttggaaacagaggtctgtcatccgaatgaagatctgaaatgtcgt	1500
Db	1441	CCAGAAATCATCTTCTTGGAAACAGGGTCTGCCATCCGATGAAGATTGAAATGTCACT	1500
OY	1501	gcacacattgtcaacaataagccccgcacagctctgtctactagtactgtgtgtgagggcaca	1560
Db	1501	GCACACATTGTCAACATAAAGCCCCGACAGCTCTGTCTACTGAGCTGTGTGAGGGCACA	1560

[illegible]

RESULT	2		
LOCUS	BC001939		
DEFINITION	BC001939 Homo sapiens, putative prostate cancer susceptibility protein, clone MGC:4102, mRNA, complete cds.	PRI	16-MAR-2001
ACCESSION	BC001939		
VERSION	BC001939.1	GI:12804972	
KEYWORDS	MGC.		

	SOURCE	human.
	ORANISM	Homo sapiens
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
	AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
	TITLE	1 (bases 1 to 2997)
	JOURNAL	Strausberg,R. Direct Submission Submitted (29-JAN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
	REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: DCTD/DTP CDNA Library Preparation: Rubin Laboratory DNA Sequencing by: The I.M.A.G.E. Consortium (LNL) BC Cancer Agency, Vancouver, BC, Canada info@cgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chin, Chris Fjell, Erin Garland, Ran Guin, Letitia Hsiao, Martin Krywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeay, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalilus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyerdunyn, Marco Marra.
	FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Series: IRAL Plate: 9 Row: 1 Column: IV. Location/Qualifiers 1..2997 /organism="Homo sapiens" /db_xref="locustid:60528" /db_xref="taxon:9606" /clone="MGC:4102 IMAGE:2820640" /tissue_type="Lung, small cell carcinoma" /clone_lib="NH_MGC-7" /lab_host="DH10B-R" /note="vector: pOTB7" 61..2541 /codon_start=1 /product="putative prostate cancer susceptibility protein" /protein_id="AAH01939.1" /db_xref="GI:12804973" /translation="MMALCSILRSACRITNSOGTTISOAPRRRPKPKDLRIHLRTRE KRGSQSGGPNVYLIQVAAGSRDSASALIVSEFERNRYLFNGCEGVQRIMQERIKLV ARLDNIFILTRHMMSNVGLSGMLITLTETGIPKCVLGSPDLKEYLAIKIFSGLLKVI IEIARPHSAPEYEDEFTWYQVFIHSEORGRKHQPOMQSERPLSRISPPSSSELN ENEBHLHGVSORGVSDSLVYAFCILHKRNFLVLAKEKGELVGRAALAPIATA AVAKCKSTIHGEPIRLIAEELCTPDDPAAPVAVVCPSDEFIOPICENATFORQKAD APVALVYHAMBPASYIVDSRIQQMERGPPTQHLVLNENCASHNLSSH10YOLNLT HPDIPLITSFRCKKEEPTLSVPWVGECCLKTYQLRRREQRAIITTCNDPEETVEA LQLENFQSOVOEYERSAQDGPAPEAKESQYPELIFTGSAIPKIRINNVATLVNISPD TSLILDGEGTFQLCRHYGDQDVLYGTAAVFVSHLADHTGLPSILOERALL ASICKPPLPLVVAPNOLKAMLOOYNHCOCEVLHHISMAKCLGEOESSPAVERL ISSILRTCPDLEEPOTCYLRICKRHAFGALVHTSSGKYVYGSDTRPCCALVRMGKDAIL LHIEATILEDGLEEAVERKTHSTSQAISVGRMMAERTIMLNHSQRAKVPILSPNFS EKGVAPADHMKVCGDFEPTMRKLIPLPKALFAGIDIEEMERRERRELDROVARALISRE LAGLEGEEPOOKKAHTEPEPOAKVKRAO"
	CDS	
	BASE COUNT	740 a 806 c 853 g 598 t
	ORIGIN	
Query Match	95.9%; Score 2379; DB 91; Length 2997;	
Best Local Similarity	99.9%; Pred. No. 0;	
Matches 2479; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	
1 atgtgggcgcttctgctgcgtcgtcgatccgcgcgcgcgaacacacattctgcagaagcgc 60		

[illegible]

QY 1141 cgcagccacagaattcaaacagctcaacctatccaccggacatcttccctctc 1200
 |||||||
 Db 1201 CGAGCCACAGATTCAAAACCACTCACTATCCACCAGACATCTTCCCTCTC 1260
 QY 1201 accagttccgcgttaagaaggagggccaccctcagttgtcccatggttcaagtgaa 1260
 |||||||
 Db 1261 ACCAGTTCCGCTTAAGAAAGAGAGGCCCACTCATGTTGCTTCAGTTCAGAGTGAA 1320
 QY 1261 tgcctcccaagtaaccagctccgltcccaaggaggttgcagagagatgcatattact 1320
 |||||||
 Db 1321 TGCCTCCTCAAGTACCAAGCTCCGTCAGAGAGAGTGGCAGAGAGATGATATTACT 1380
 QY 1321 tgcgaatccaggaattcaagtgttggagctcaggttcccaattccagcagagctg 1380
 |||||||
 Db 1381 TGCATCTCTAGGAATTCATAGTTGAGGCGCTCTCAAGCTTCCAACTTCACAGACGGTG 1440
 QY 1381 caggaatcacagagaggtgagcagagagccagcccaagcagagaaagaatcagttac 1440
 |||||||
 Db 1441 CAGAGTACAGAGAGAGTGGCAGAGAGGCCACAGCCCAAGAGAAAGATCACTAC 1500
 QY 1441 ccagaaatcaatcttcccttggacaagggtctgcattcccgatgaattcgaatlytca 1500
 |||||||
 Db 1501 CCAGAAATCATCTTCTTGGAGACAGGGCTGCAATCCCGATGAAGATTGCAATGTCAGT 1560
 QY 1501 gccaacattgcaacataagcccgagagctctcgtactgactgagtggtgagggaca 1560
 |||||||
 Db 1561 GCCACACTTCTCAACATTAACCCCGACACGCTCTCTACTGTGAGTGTGAGGGCAG 1620
 QY 1561 ttctggagcagctgtgcctcaatcagagacaggttgagacaggttccctggacacctgct 1620
 |||||||
 Db 1621 TTGGGAGAGTGTGCGCTCATTTACGGAGACAGAGTGGACAGGGTCCCTGGGACACCTGCT 1680
 QY 1621 gctgtgtgtgtgtcccaactgcagcagatcacacaaggggttgcgaatctctgtc 1680
 |||||||
 Db 1681 GCTGTGTGTGTGCCACCGCAGCAGATCACACAGGGGCTTGCAAGATCTTCTGCTG 1740
 QY 1681 cagaagagaaagcgccttggatcttgggaaagcgccttcaaccttgcgtgtgtgtgc 1740
 |||||||
 Db 1741 CAGAGAAAGCGCCCTGGGATCTTTGGAAACCGGCTTCAACCTTGGCTGTGGTGGCC 1800
 QY 1741 cccaacagctcaaaagcctgtcccaagcagttacacaacaggtgcagaggtctctcac 1800
 |||||||
 Db 1801 CCCAACCAAGCTCAAGGCTGGCTCCAGACATACCACAAACAGTCCAGAGGTCCTGAC 1860
 QY 1801 cacatcagtaattccctgcgcaaatgccttcaggaagggtcagatccagctcgtgca 1860
 |||||||
 Db 1861 CACATCAGTATGATTCCTGCAATGCTTCAAGGAAGGGCTGAGATCTCCAGTCTGCA 1920
 QY 1861 gttgaaagatgtatcagttcgtgtgtgcgaacatgtgatttgaagaagttcagaacct 1920
 |||||||
 Db 1921 GTGGAAAGATTGATTCATGCTGTGGCAACATGTAATTTGGAAGAGTTTCAACACTGT 1980
 QY 1921 ctgtgagcagcagcaagcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1980
 |||||||
 Db 1981 CTGGTGGGAGCTGCAAGCATGCGTTTGGCTGTGCGCTGTGTCACACTGTGCTGAAA 2040
 QY 1981 gttgtctatctcgggagacacatgcccctgcagagctctgtgtgtgtgtgtgtgtgt 2040
 |||||||
 Db 2041 GTGGTCTATTCGGGGGACACATGCTGCGAGGCTCTGTGTCGGATGGGAAAGATGCC 2100
 QY 2041 accctcctgtaataagaagcaccctgtgaaagatgttgaagaaggagcaggtgaaag 2100
 |||||||
 Db 2101 ACCCTCTGTATGATGAAGCCACCTGGAAGATGTTGGAAGGAGAGTGTGGAAG 2160
 QY 2101 acacacagcacaagctcccaagcagcatcagctgtggaatgagagcagagttcat 2160
 |||||||
 Db 2161 ACACACAGCACAAGCTCCCAAGCATCAGCGTGGGATGGCGATGAACCGCGAGTTCA 2220
 QY 2161 atgtcgaacactcagcagcagctatgtcgaaggttccctctcagcccnaactcagc 2220
 |||||||
 Db 2221 ATGCTGAACACTTACAGCAGCGCTATGCAAGGTCCCTCTTCAAGCCCAACTTACG 2280

QY 2221 gagaagttggagttgtgcttgaccacatagaagttcgtcttggagaattccacaatg 2280
 |||||||
 Db 2281 GAGAAATGGGAGTTGCTTTGACACATGAAGTCTGCTTTGGAGACTTTCCAAACATG 2340
 QY 2281 cccaagctgtatcccccacatgaagccctgttgtgtgcacalcagagaatgagagag 2340
 |||||||
 Db 2341 CCCAAGCTGATTCGCCACAGTGAAGCCCTGTTGCTGCGACATCGAGAGATGGAGAG 2400
 QY 2341 cgcagagagaagcggagagctgcgcaaggtgcgagcggcctctgtccagagagctgca 2400
 |||||||
 Db 2401 CGCAGGAGAGAGCGGGAGCTGCGCAGAGTCCGGCGGCTCTCTGTCCAGGAGCTGGCA 2460
 QY 2401 ggcgacctgagagatggagagcctcagcagaagcgggcccacagagagagcacaagcc 2460
 |||||||
 Db 2461 GCGGCGCTGAGAGTGGGAGCTTACAGAGAGGGGCCACACAGAGGCCACAGGCC 2520
 QY 2461 aagaaggtcagagccacagtgca 2481
 |||||||
 Db 2521 AAGAGGTACAGGCCAGTGA 2541

RESULT 3
 BC004158
 LOCUS
 DEFINITION
 Homo sapiens, putative prostate cancer susceptibility protein,
 clone MGC:2441, mRNA, complete cds.
 ACCESSION
 BC004158
 VERSION
 BC004158.1 GI:13278770
 KEYWORDS
 MGC.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 1 (bases 1 to 3006)
 AUTHORS
 Strausberg, R.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (01-MAR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
 USA

REMARK
 COMMENT
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
 Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>

contact: amadan@systemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
 Greene, Mark Keltman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 2 Row: C Column: 13.

FEATURES
 source
 Location/Qualifiers

1..3006
 /organism="Homo sapiens"
 /db_xref="locusid:60528"
 /db_xref="taxon:9606"
 /clone="MGC:2441 IMAGE:2820640"
 /tissue_type="lung, small cell carcinoma"
 /clone_lib="NIH_MGC_7"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 70..2550
 /codon_start=1
 /product="putative prostate cancer susceptibility protein"
 /protein_id="AAH04158.1"
 /db_xref="GI:13278771"
 /translation="MVALCSILRSAGRTMSOGRITISQAPARRPRKDPRLRLPTRE
 KRPSGSGSPNIVYLQVVAAGSRDSGALYVSEENRYLFLNCGEVQRLMQEHLKLV

Db	1930	GTGGAAAGATTGATCAGTTCGGCTTTGGCGAAGCATGTATTTGGAGAGCTTTCAGACCTGT	1989
Qy	1921	ctggctgcggcaatctgcaagcatctgtctgtgctgtgcgtctggtgtcacacatcttgctgtgaa	1980
Db	1990	CTGGTGGGGGCATCTCAGAGCATGCTTTTGGGCTGTGCGCTGGTGGACACCTCTGGGCTGGAAA	2049
Qy	1981	gtgtgtctattccgggagacacatgacctgtcgaggtcttgcttgatccgatatgggaaagatgccc	2040
Db	2050	GTGGTCTATTTCGGGGGACACCATCTCCCTGGAGAGCTCTGTGCTCGGATGGGGAAGATGGCC	2109
Qy	2041	acctctctgatacatatgagaccaccccttgaaagatgattcttgaaagagaaagaaatggaag	2100
Db	2110	ACCTCTCTGATACATGAAAGCCACCCCTGGAAGATGTGTTGCAAGAGCAAGCATGTGGAAG	2169
Qy	2101	acacacagcacaaagcttcccaagccatcaagcgttgggagatgcygatatgaacgcgagattcat	2160
Db	2170	ACACACAGCACAAAGCTCCCAAGCATGAGCGTGGGATGGGATGAAGACGGGAGTTGATT	2229
Qy	2161	atgtctgaaccacatctcagtcagcgtatgcaaaaggttccccctctctaaagcccaactcagc	2220
Db	2230	ATGTGTGAACCTTCCTCACACAGCGGTATGCCAAAGTCTCCCTCTTCACACCCCAACTTCAGC	2289
Qy	2221	gagaagaatgggagatgtgctctttgacacatatgaagtctgctctttgagagcttccaacaatg	2280
Db	2290	GAGAAAGTGGGAGATTGCCCTTTGACCAATGAAAGTCTTGCTTTGGAGACTTTCCAAACATG	2349
Qy	2281	cccaagctgtatccccccacatgaaagccctgttctgtgcgacatcagagagatgagagag	2340
Db	2350	CCCAAGCTGATTCCTCCCACTGAAGCCCTGTCTCTGTGGCAGCATCGAGAGATGGAGAG	2409
Qy	2341	cgcaaggagaagcgggagctgtgcgcaaggtgcggcgagccctctgtctcaaggagctgtgca	2400
Db	2410	CGCAGGGGAGAGCCGGAGCTCTGGCAGAGTGGCGGGCGCCCTCTCTGTCCAGGAGCTGGCA	2469
Qy	2401	ggcgccctgggaggttggggagccctcagaagaaggggccccacaaggagggccaacgggcc	2460
Db	2470	GGCGCCCTGGAGATGGGGAGCCCTCAGCAGAAAGCGGGCCCAACAGAGGAGCCACAGGCC	2529
Qy	2461	aagaaggtcagaagcccaagtga	2481
Db	2530	AAGAAGGTCAAGGCCCACTGA	2550
RESULT 4			
AK001392			
LOCUS	2976 bp	mRNA	PR1 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10530 fls, clone NT2RNP2000985.		
ACCESSION	AK001392		
VERSION	AK001392.1 GI:7022621		
KEYWORDS	oligo capping; fls (full insert sequence).		
SOURCE	Homo sapiens teratocarcinoma cell line:NT2 cDNA to mRNA, clone:1fb:NT2R2 clone:NT2R2P2000985.		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	1 (sites)		
	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiraatori,A., Sudo,H., Wagaatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Morakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakatsuki,A., Nakamura,Y., Nagahari,K., Masuhio,Y., Ninomiya,K. and Iwayanagi,T.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished (2000)		
REFERENCE	2 (bases 1 to 2976)		
AUTHORS	Isogai,T. and Otsuki,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao		
	Isogai, Heli Research Institute, Genomics Laboratory, 1532-3 Yanai		
	Kisarazu, Chiba 292-0812, Japan (E-mail:genomices@hri.co.jp,		
	tel:81-438-52-3951, fax:81-438-52-3952)		

[illegible]

OY	421	octccacacacitgagaaataataccttgaaagcaatacaaatattctctgltccattgaaagga	480
Db	495	cctccacacacitgagaaataatacttgcgaagcaatcaataattttctgtccattgaaagga	554
OY	481	atagacatgctgtgctgagccaccctctgcccagaatacaggaatgaacacatgacgt	540
Db	555	atagaaactgacctatgcccccccaactctgcccagaatagcagaaatgaacacatgacgt	614
OY	541	taccagatcccaatacaacgtgaacagaagagggaacacacacatgacagatgcca	600
Db	615	taccagatcccccaatacaacaggaacagagaggggaaagaccacacacatgacagatgcca	674
OY	601	gaaagacctctcaagaagctcaagtcagagcagatcttaagacatccgaatcgaatgaatat	660
Db	675	gaaagacctctcaagaagctcaagtcagagcagatcttaagacatccgaatcgaatgaatat	734
OY	661	gagccacacacttccacatacgtggtatgacagagaagagggtcagaggaactctccctgctc	720
Db	735	gagccacacacttccacatacgtggtatgacagagaagagggtcagaggaactctccctgctc	794
OY	721	gtagcttccatctgttaagcttcaacttaagaagaggaaactctctgtgtcctaaagcaag	780
Db	795	gtagcttccatctgttaagcttcaacttaagaagaggaaactctctgtgtcctaaagcaag	854
OY	781	gagatgagcccccacagatctggagacagctgacatgcctcccaatactgctctctcaagagac	840
Db	855	gagatgagcccccacagatctggagacagctgacatgcctcccaatactgctctctcaagagac	914
OY	841	gggaaagaatcaatcatatgaagaagaagatcttgctgtaagaagctgtgtacccctcca	900
Db	915	gggaaagaatcaatcatatgaagaagaagatcttgctgtaagaagctgtgtacccctcca	974
OY	901	gacctgtgtgtgtcttctgtgtgtgtagaatgtccagatagaagactcaatacaacatc	960
Db	975	gacctgtgtgtgtcttctgtgtgtgtagaatgtccagatagaagactcaatacaacatc	1034
OY	961	tgtgaagaatgcacacttccagaaagtaccaaaggaagcagatgccccgttgacctgtgtg	1020
Db	1035	tgtgaagaatgcacacttccagaaagtaccaaaggaagcagatgccccgttgacctgtgtg	1094
OY	1021	gttcaacatgagcccggaagatctgtgctgtgtagaagaagtacagaaagtgtgtgaagag	1088
Db	1095	gttcaacatgagcccggaagatctgtgctgtgtagaagaagtacagaaagtgtgtgaagag	1154
OY	1081	tttgagccctgcacccagacacttgctccctaagtgaagaaactgtgcctcagttcaacaact	1140
Db	1155	tttgagccctgcacccagacacttgctccctaagtgaagaaactgtgcctcagttcaacaact	1214
OY	1141	cgcagccaagaattcaaaacccagctcaacatcaataccacggaacatcttcccctgctc	1200
Db	1215	cgcagccaagaattcaaaacccagctcaacatcaataccacggaacatcttcccctgctc	1274
OY	1201	acacagtttcgcgtgtaagaagaggagggccacacctaagtgtgcccattgttaagggtga	1266
Db	1275	acacagtttcgcgtgtaagaagaggagggccacacctaagtgtgcccattgttaagggtga	1334
OY	1261	tgcctcctcaagtaecagctccgtccacagaagaggatgtgcagagaggtatgcataatact	1320
Db	1332	tgcctcctcaagtaecagctccgtccacagaagaggatgtgcagagaggtatgcataatact	1391
OY	1321	tgcacatcctgaggaattcatatgatttgagggcgtctgcagctcccaacttccagcaagggctg	1388
Db	1392	tgcacatcctgaggaattcatatgatttgagggcgtctgcagctcccaacttccagcaagggctg	1451
OY	1381	caggagatcaagagaggatgtgcagagagggccacagccacagacagagaagaagaatcaatgac	1440
Db	1452	caggagatcaagagaggatgtgcagagagggccacagccacagacagagaagaagaatcaatgac	1511
OY	1441	ccagaaatcatcttcccttgtaacaagaggtctgcatacccgatgaagaatctcgaatgtcagtt	1500
Db	1512	ccagaaatcatcttcccttgtaacaagaggtctgcatacccgatgaagaatctcgaatgtcagtt	1571

QY	1501	gceaeacttgcacaacaagcccgaaagctctctctactctgactctgctgagagacaca	1500
Db	1572	gccacacttgcacacatpaaccccgaaacctctctctactgactgctgagagacaca	1631
QY	1561	tttgagagctctgctgcctactaagagacagctggaagagtcctctgagacccctgct	1620
Db	1632	tttgaggcagctgctgcctctactaagagacagctggaagagtcctctgagacccctgct	1691
QY	1621	gctgtgttctgtgccacactgcaagacatcacacacagagcttgcaagatcttctgtg	1680
Db	1692	gctgtgttgtgttcctccactgcacagacagatcacacacagagcttgcaagatcttctgtg	1751
QY	1681	cagagagaagcgccctctggcatctttgggaagcgcttcaaccttgcctgtgtgttgc	1740
Db	1752	cagagagaagccgccccttggcatctttgggaagcccttccaccttgcctgtgtgttgc	1811
QY	1741	cccaacacagctcaagactctgcttccagcaatcacacacagctgcagagagcttctgac	1800
Db	1812	cccaacacagctcaagactctgcttccagcaatcacacacagctgcagagagcttctgac	1871
QY	1801	cacatcagatgatcttccctgcacaatgcttcaagaaagggctgaatctcagctctgca	1860
Db	1872	cacatcagatgatcttccctgcacaatgcttcaagaaagggctgaatctcagctctgca	1931
QY	1861	gtggaagatgatcagctctgcctgttgcgaacatgtgatttggaagagttccaagctgt	1920
Db	1932	gtggaagatgatcagctctgcctgttgcgaacatgtgatttggaagagttccaagctgt	1991
QY	1921	ctgtgtgcgacctcgaagcatgcttgcgtgtgcgtgcgtgcgtgcgtgcgtgcgtgcgt	1980
Db	1992	ctgtgtgcgacctcgaagcatgcttgcgtgtgcgtgcgtgcgtgcgtgcgtgcgtgcgt	2051
QY	1981	gtgtgtctattcccgagacacacatgcccctcgaagctctgttgcgtatgtaggaatgccc	2040
Db	2052	gtgtgtctattcccgagacacacatgcccctcgaagctctgttgcgtatgtaggaatgccc	2111
QY	2041	acctctctgatacatgaagccaccttgaagatgcttggaaaggaagcagtggaagaa	2100
Db	2112	acctctctgatacatgaagccaccttgaagatgcttggaaaggaagcagtggaagaa	2171
QY	2101	aacacagacaaagctcccaagacatcaagctgtaggaatgtaggaatgtaggaatgtagga	2160
Db	2172	aacacagacaaagctcccaagacatcaagctgtaggaatgtaggaatgtaggaatgtagga	2231
QY	2161	atctgtgaacccactcagcagcagctatgccaagtgctccccccttcaagcccaactcagc	2220
Db	2232	atctgtgaacccactcagcagcagctatgccaagtgctccccccttcaagcccaactcagc	2291
QY	2221	gagaaagtggagtgctgcttctgaacacatgaagtgctgcttctgaagcttccaacaatg	2280
Db	2292	gagaaagtggagtgctgcttctgaacacatgaagtgctgcttctgaagcttccaacaatg	2351
QY	2281	cccaagctgattccccaactgaagaccctgttgcctgcggaacatcgaagagatggaagag	2340
Db	2352	cccaagctgattccccaactgaagaccctgttgcctgcggaacatcgaagagatggaagag	2411
QY	2341	cgcagaggaagcgcgagctgcgagcagctgtagcgcgcctcctgttccaggaagctgcga	2400
Db	2412	cgcagaggaagcgcgagctgcgagcagctgtagcgcgcctcctgttccaggaagctgcga	2471
QY	2401	ggcgcgcctggaagatggg 2418	
Db	2472	ggcgcgcctggaagatggg 2489	

RESULT 5 AF308698 2908 bp mRNA 27-FEB-2001
LOCUS Pan troglodytes ELAC2 mRNA, complete cds.
DEFINITION AF308698
ACCESSION AF308698.1 GI:10946496
KEYWORDS chimpanzee.
SOURCE

QY 1141 ccgagcccaagaatcacaaccagctcaacctcaccacccggacatctccccctgctc 1200
|||||
Db 1141 cgcagccacaaagatcaaacccagctcaacctcaccacccggacatctccccctgctc 1200
QY 1201 accagcttcgcgtgtagaagaaggccccaccctcagtgctgccaatggttcaggatga 1260
|||||
Db 1201 accagcttcgcgtgtagaagaaggccccaccctcagtgctgccaatggttcaggatga 1260
QY 1261 tgcctcctcaagtaacagctcgcgtcccaaggaggagtgagcagaagagtgccatctact 1320
|||||
Db 1261 tgcctcctcaagtaacagctcgcgtcccaaggaggagtgagcagaagagtgccatctact 1320
QY 1321 tgcgaatccctgaagaaatcctagttgagcgctgcaagcttcccaactccagcagaagcgtg 1380
|||||
Db 1321 tgcgaatccctgaagaaatcctagttgagcgctgcaagcttcccaactccagcagaagcgtg 1380
QY 1381 cagaggaatcagaaggagtgctgcgagagcgccccagcccaagaagaagaagtaagtaagtaac 1440
|||||
Db 1381 cagaggaatcagaaggagtgctgcgagagcgccccagcccaagaagaagaagtaagtaagtaac 1440
QY 1441 ccagaaatcatctctcctctggaacaaggctctgcacatccgaatgagaatctgcaatgcaagt 1500
|||||
Db 1441 ccagaaatcatctctcctctggaacaaggctctgcacatccgaatgagaatctgcaatgcaagt 1500
QY 1501 gccacacttgtaacaataagccccagacagctctgctactgactgagctggtgtagaggacaca 1560
|||||
Db 1501 gccacacttgtaacaataagccccagacagctctgctactgactgagctggtgtagaggacaca 1560
QY 1561 ttggggcagactgctggcgctgtaattatgaggaacagagtgtagaaggagtgctgggacacctgact 1620
|||||
Db 1561 ttggggcagactgctggcgctgtaattatgaggaacagagtgtagaaggagtgctgggacacctgact 1620
QY 1621 gctggtgtgtgtgtcccaactgcaagcagagatcacaacagcaggtctgccaagtatctgtctg 1680
|||||
Db 1621 gctggtgtgtgtgtcccaactgcaagcagagatcacaacagcaggtctgccaagtatctgtctg 1680
QY 1681 cagagagaacgcgcctctgagatcttgggaaagcgcttcaaccttgcctgctggtgtgtgccc 1740
|||||
Db 1681 cagagagaacgcgcctctgagatcttgggaaagcgcttcaaccttgcctgctggtgtgtgccc 1740
QY 1741 cccaaccagctcaaaagcctggtctccagcaggtacacaacacagtgccaagaagtgctccgcac 1800
|||||
Db 1741 cccaaccagctcaaaagcctggtctccagcaggtacacaacacagtgccaagaagtgctccgcac 1800
QY 1801 cacatcagatgatatctcctgccaatgctctcaggaagggtctgagatctcccaatctctgca 1860
|||||
Db 1801 cacatcagatgatatctcctgccaatgctctcaggaagggtctgagatctcccaatctctgca 1860
QY 1861 gtcggaagaatctgacagctgctgctgtgtagaacaatgtaattggaagaagttcagacactgt 1920
|||||
Db 1861 gtcggaagaatctgacagctgctgctgtgtagaacaatgtaattggaagaagttcagacactgt 1920
QY 1921 ctggtgagcagcactgcaagaagatgctgtgctgtgctgtagcacaacctctgctgtgaa 1980
|||||
Db 1921 ctggtgagcagcactgcaagaagatgctgtgctgtgctgtagcacaacctctgctgtgaa 1980
QY 1981 gtcgtctatctcgggagaaaccatgccccctcgagaggtcctgctcggagtagggaaagatgccc 2040
|||||
Db 1981 gtcgtctatctcgggagaaaccatgccccctcgagaggtcctgctcggagtagggaaagatgccc 2040
QY 2041 accctctcgtatcatatgaagccacacctggaagaatggtttggaagaagaagcagtggaaga 2100
|||||
Db 2041 accctctcgtatcatatgaagccacacctggaagaatggtttggaagaagaagcagtggaaga 2100
QY 2101 acaacagcagaacaagctcccaagccatcagcgtggtggaatgagatgagcggagatcatatt 2160
|||||
Db 2101 acaacagcagaacaagctcccaagccatcagcgtggtggaatgagatgagcggagatcatatt 2160
QY 2161 atgctgaacacattcagcagcagcgtatgcaaggttccccctctcagcccccaactcagc 2220
|||||
Db 2161 atgctgaacacattcagcagcagcgtatgcaaggttccccctctcagcccccaactcagc 2220

QY 2221 gagaagaatggaagtgctgctcttgaccacaatgaagtgctgcttggagagacttccacaatg 2280
|||||
Db 2221 gagaagaatggaagtgctgctcttgaccacaatgaagtgctgcttggagagacttccacaatg 2280
QY 2281 cccaagctatctcccccacttgaaagccctgcttgcctgagagacatgagatgagagag 2340
|||||
Db 2281 cccaagctatctcccccacttgaaagccctgcttgcctgagagacatgagatgagagag 2340
QY 2341 cgcagggagaagaagcggagagctgcgcgagtgctggcgagccctctctcagaggaagtgca 2400
|||||
Db 2341 cgcagggagaagaagcggagagctgcgcgagtgctggcgagccctctctcagaggaagtgca 2400
QY 2401 ggcgccttggaagatgaggagcctcagcagaa 2432
|||||
Db 2401 ggcgccttggaagatgaggagcctcagcagaa 2432
RESULT 6
AF308694 2893 bp mRNA PRI 27-FEB-2001
LOCUS
DEFINITION
Gorilla gorilla ELAC2 (ELAC2) mRNA, complete cds.
ACCESSION
AF308694.1 GI:10946488
VERSION
KEYWORDS
SOURCE
ORGANISM
gorilla.
Gorilla gorilla
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Gorilla.
REFERENCE
1 (bases 1 to 2893)
Tavligian, S.V., Simard, J., Teng, D.H.F., Abtin, V., Baumgard, M.,
Beck, A., Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M.,
Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C.,
Ghaifari, S., Gupta, J.S., Hu, R., Iliev, D., Janekci, T., Kort, E.N.,
Laliv, K.E., Leavitt, A., Leblanc, G., McArthur-Morrison, J.,
Pederson, A., Penn, B., Peterson, K.T., Reid, J.E., Richards, S.,
Schroeder, M., Smith, R., Snyder, S.C., Swedlund, B., Swensen, J.,
Thomas, A., Tranchant, M., Woodland, A.M., Labrie, F., Skolnick, M.H.,
Neuhansen, S., Rommens, J., and Cannon-Albright, L.A.
A candidate prostate cancer susceptibility gene at chromosome 17p
11175785
Nat. Genet. 27 (2), 172-180 (2001)
TITLE
JOURNAL
PUBMED
11175785
REFERENCE
AUTHORS
2 (bases 1 to 2893)
Tavligian, S.V., Simard, J., Teng, D.H.F., Baumgard, M., Beck, A.,
Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M.,
Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaifari, S.,
Gupte, J.S., Hu, R., Iliev, D., Janekci, T., Kort, E.N., Laliv, K.E.,
Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B.,
Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R.,
Snyder, S.C., Swedlund, B., Swensen, J., Thomas, A., Tranchant, M.,
Woodland, A.-M., Labrie, F., Skolnick, M.H., Neuhansen, S., Rommens, J.,
and Cannon-Albright, L.A.
Gorilla gorilla ortholog of human Hpc2/ELAC2
TITLE
JOURNAL
PUBMED
Unpublished
REFERENCE
AUTHORS
3 (bases 1 to 2893)
Tavligian, S.V., Simard, J., Teng, D.H.F., Baumgard, M., Beck, A.,
Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M.,
Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaifari, S.,
Gupte, J.S., Hu, R., Iliev, D., Janekci, T., Kort, E.N., Laliv, K.E.,
Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B.,
Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R.,
Snyder, S.C., Swedlund, B., Swensen, J., Thomas, A., Tranchant, M.,
Woodland, A.-M., Labrie, F., Skolnick, M.H., Neuhansen, S., Rommens, J.,
and Cannon-Albright, L.A.
Direct submission
TITLE
JOURNAL
Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
Lake City, UT 84103, USA
FEATURES
source
location/Qualifiers
1..2893
/organism="Gorilla gorilla"
/db_xref="taxon:9593"
1..2893
/gene="ELAC2"
1..2481
CDS

Db	1741	CCGAGCGACCTCAAAAGCCCTGGGCTCCAGCAGTACCAACACCGTCCAGGAGGTCTGCAC	1800
QY	1801	cacatcagtaataatcttcctcgcacaatgctctcaggaaaggctctgagatctctccagtcctga	1860
Db	1801	CACATCAGTATATTCCTCCCAATCCCTTCAGGAAGGGGGCTGAGATCTCCAGTCTCTGA	1860
QY	1861	gtcggaagaatgatcagcttcgcctcgtttgcgaacaatgtaatttgaaagtctcagactgt	1920
Db	1861	GTGGAAGAAGTATCATCGTTCCGCTGTTGCCAACAATGTGATTGGACAGATTTCAGACCTGT	1920
QY	1921	ctggtgcggcactgcgaagcatgctgtttggcgtctgcgtcgtgcgaacactctggtcgtgaaa	1980
Db	1921	CTGTGTGGCGCACTGCACAGCATCGTTGGTGGCTGTGCCCTGTGCACACCTGTGGCTGGAAA	1980
QY	1981	gtcgatctatcccggggaacacatgcccctcggaggctctgctcggatgggggaagaatgac	2040
Db	1981	GTGATCTATTCGGGGACACCATGCCCCCTGCGAGCGCTGTGTCCGCATGGGGAAAGATGCC	2040
QY	2041	acctctcgtatacatgaaagccaccccttgaaagatggtttggaagaggaaacagctggaagaag	2100
Db	2041	ACCCCTCTGATACATGAACCCACCCCTGGAAAGTGTGTTGGAAGAGAAACACAGTGGAAAG	2100
QY	2101	aacacagagaacaggtcccaagaacatcaacgtggggaatcgatatgaacggcgaagtcat	2160
Db	2101	AACACAGACACAACTGCTCCCAAGCCATACACGTGGGATCGGATGACAGGGAGTTCAAT	2160
QY	2161	atctcgtaacactctcaagcagcgtatgccaaggtcccccctctccaccccaactctcagc	2220
Db	2161	ATCTGTGAACCACTTCAAGCAGCGCTATGCTCAAGAGTCCCTCTTCAGCCCCAACTTCAAC	2220
QY	2221	ggaagaatggaggttgaccttgacacacatgaaaggtctggtttggagaacttccaagaatg	2280
Db	2221	GGAAGAATGGAGTTGCCCTTTGACCAATGAAGGTCTGCTTTGGAACTTCCAAACAATG	2280
QY	2281	cccaagctgattccccccaatgaaagccctgttttgcctcgtcgaacatcgaaagagatggaagag	2340
Db	2281	CCCAAGCTATTCCTCCCACTGAAAGCCCTGTTTGCCTCCGGCAGATCGAAGAGATGAGAGAG	2340
QY	2341	cgcaaggagaaacggaagctctgcgcagagtgcggggcgagccctcctctc	2388
Db	2341	CGCAGGGAACGCGGAGCTGCGGCGGAGTGGCGGCGGCGGCTCTCTCTGACC	2388

RESULT	7
LOCUS	AC005277
ACCESSION	AC005277
VERSION	AC005277.1
KEYWORDS	GTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 (bases 1 to 118788)
TITLE	Bittern,B., Fasmann,K., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL	Homo sapiens chromosome 17, clone hRRK.597_M_12
REFERENCE	unpublished
AUTHORS	2 (bases 1 to 118788)
	Bittern,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Bonth,C., Boutwell,C., Brown,A., Castle,A., Cenny,J., Cooke,P., Depayre,E., Devon,K., Dewar,K., Doneelan,L., Etemadi,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S., Geraiery,K., Gilmartin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L., Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K., Melidim,J., Molla,M., Morris'W., Morrow,J., Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Stickland,C., Subramanian,A., Tortorella-Miller,I., Vassiliev,H.,

[illegible]


```
unsure      25816.    25895
            /note="Single-stranded coverage."
repeat_region complement(25994. ,26099)
            /rpt_family="purine-rich"
repeat_region complement(26122. ,26291)
            /rpt_family="MLTID"
                26312.    26337
repeat_region /rpt_family="(CAAAA)n"
                26456.    26480
repeat_region /rpt_family="(GGGAA)n"
                complement(26576. ,26781)
repeat_region /rpt_family="MER6B"
                complement(26932. ,26990)
repeat_region /rpt_family="MER21B"
                26991.    27290
repeat_region /rpt_family="Alusx"
                complement(27291. ,27975)
repeat_region /rpt_family="MER21B"
                28034.    28197
repeat_region /rpt_family="LNR28"
                28266.    28795
repeat_region /rpt_family="LNR28"
                28843.    28916
repeat_region /rpt_family="LNR28"
                28920.    29005
repeat_region /rpt_family="MTR"
                29140.    29233
```

Query Match 10.0%; Score 247; DB 85; length 118788;
Best Local Similarity 100.0%; Pred. No. 2.6e-131;
Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

RESULT	8
AF304371.S2	
LOCUS	AF304371.s2 740 bp DNA PRI 19-OCT-2000
DEFINITION	Homo sapiens putative prostate cancer susceptibility protein
ACCESSION	HPc2/ELAC2 gene, exon 17 and partial cds.
VERSION	AF304369.1 GI:10880929
KEYWORDS	.
SEGMENT	2 of 2
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 740)
AUTHORS	Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carrillo,A.R., Chen,Y., Dayananth,P., Desrochers,M., Dunont,C.M., Farnham,J.M., Frank,D., Frye,C., Gaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,

Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J., and Cannon-Albright,L.A.
A strong candidate prostate cancer susceptibility gene at chromosome 17p

TITLE
JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 740)
Tavligian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayanath,P., Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S., Gupta,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Lally,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J., and Cannon-Albright,L.A.
Direct Submission
Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt Lake City, UT 84108, USA

TITLE
JOURNAL
FEATURES
source

1. 740
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p"
/number=17
301. .439
<301. .>439
/product="putative prostate cancer susceptibility protein
HPC2/ELAC2"
<301. .>439
/codon_start=2
/product="putative prostate cancer susceptibility protein
HPC2/ELAC2"
/protein_id="AAG24440.1"
/db_xref="GI:10880931"
/translation="PDTSLIDCGEGTFCQLCRHYGDQYRVLTGLAAYVSHLADH
HP"

BASE COUNT 154 a 188 c 173 g 225 t

ORIGIN

Query Match 5.7%; Score 142; DB 89; Length 740;
Best Local Similarity 100.0%; Pred. No. 2.5e-70;
Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1519 agcccccgaacagctctgtctgactgagctgtgtgagggcacatttgagcgtgtgcgt 1578
|||||
Db 299 agcccccgaacagctctgtctgactgagctgtgtgagggcacatttgagcgtgtgcgt 358
|||||

QY 1579 cattacgagacacaggttgacagaggtcctgtgagccctgtgtgtttgtgtccac 1638
|||||
Db 359 cattacgagacacaggttgacagaggtcctgtgagccctgtgtgtttgtgtccac 418
|||||

QY 1639 ctgacgcagatcacacacgg 1660
|||||
Db 419 ctgacgcagatcacacacgg 440
|||||

RESULT 9
AF304371S1 721 bp DNA PRI 19-OCT-2000
LOCUS Homo sapiens putative prostate cancer susceptibility protein
DEFINITION HPC2/ELAC2 gene, exon 7 and partial cds.
ACCESSION AF304371
VERSION AF304371.1 GI:10880928
KEYWORDS
SEGMENTS 1 of 2
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 721)

Tavligian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayanath,P., Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S., Gupta,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Lally,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J., and Cannon-Albright,L.A.
A strong candidate prostate cancer susceptibility gene at chromosome 17p

TITLE
JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 721)
Tavligian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayanath,P., Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S., Gupta,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Lally,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J., and Cannon-Albright,L.A.
Direct Submission
Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt Lake City, UT 84108, USA

TITLE
JOURNAL
FEATURES
source

1. 721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p"
/number=7
301. .420
<301. .>420
/product="putative prostate cancer susceptibility protein
HPC2/ELAC2"
<301. .>420
/codon_start=3
/product="putative prostate cancer susceptibility protein
HPC2/ELAC2"
/protein_id="AAG24439.1"
/db_xref="GI:10880930"
/translation="EORRGKHQWQSPERPLSRSLSPERSDSSESNENEPHPH"

BASE COUNT 201 a 143 c 184 g 193 t

ORIGIN

Query Match 4.9%; Score 122; DB 89; Length 721;
Best Local Similarity 100.0%; Pred. No. 1.1e-58;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 560 gtgaaagagagagagaaagcaccacacatgagaggtccagaagaagccttcagcagc 619
|||||
Db 301 gtgaaagagagagagaaagcaccacacatgagaggtccagaagaagccttcagcagc 360
|||||

QY 620 tcagtcacagcagatctcagactccagatcgaaatgaaatgacacaccttcacatg 679
|||||
Db 361 tcagtcacagcagatctcagactccagatcgaaatgaaatgacacaccttcacatg 420
|||||

QY 680 gt 681
||
Db 421 gt 422

RESULT 10
AF308696 2712 bp mRNA ROD 27-FEB-2001
LOCUS Mus musculus ELAC2 mRNA, complete cds.
DEFINITION AF308696
ACCESSION AF308696
VERSION AF308696.2 GI:11992378
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus

16609..16667,16746..16818,17370..17482,18671..18766,
21539..21677,21757..21842,23553..23671,24619..24715,
27291..27429,27763..27801,27930..28039,28552..28652,
28733..28853,29101..29179,29710..29854,30406..30624,
31242..31277)
/gene="Ela2"
/note="alternatively spliced 25 exon form"
/product="putative prostate cancer susceptibility protein"
<470..>31277
/gene="Ela2"
join(8470..8690,9077..9127,9266..9336,9948..10012,
10957..11014,12096..12164,12933..13064,14902..14960,
16609..16667,16746..16818,17370..17482,18671..18766,
21539..21677,21757..21842,23553..23671,24619..24715,
27291..27429,27763..27801,27930..28039,28552..28652,
28733..28853,29101..29179,29710..29854,30406..30624,
31242..31277)
/gene="Ela2"
/note="25 exon form"
/codon_start=1
/product="putative prostate cancer susceptibility protein"
/protein_id="AAK29420.1"
/db_xref="GI:13540342"
/translation="MVALRSILRPLGLRTMSGSRARPPKDPRLHRLTRKRGPG
GGPNTVYLVQVVAAGRDAGALYFSEVNRVYFNGCEGQRLMOEHLKVALDNIPL
TMMHSMVGLCGMILTLTKETGLPKCVLSGPQLKYLBAIKTFSGPLKGLVALNPH
SAPETKDEMTYVQVPIHSEKRGKQSPSPRTSPNRLSPKSSDSGAENGQCPPE
DSSAGANRAWRAGDSLVAVFCKLHLKRGNFVLKAKELGPGTAAAPADIAVAK
GSIYEGREIAEELCTPPDGLVFIYECDEGFIPLICENDTFKRAQADAPVA
LVVHIAPESVLIDSRVQOMEREPDTHLLENKPSVHNRSKIOQLDILHPI
FQULSFYSKEGSLVPTVTRGECILKQLPRKRWMDTLDNCSLTVNLAELELP
SFQESVEEYKRWQENPAPAEKRSQYPELVIGTSALFMKIRNVSITLVNLSPKSV
LIDGEGFTGQCRHGOIDRVCLSLTAVFSHLHADHTGLNLILQREHALSL
KFPQPLVVAFTQRLAMLDQYHNHCOEILHNVSMIPAKLOGAESNTTLELISL
LETCDLEFQCTLVHCKHAFGALVHSSGMVYSGDTPCEALVOMKQADTLIHE
ATLEDGEEAEVKTSTSOAINVGMNRNAEFIMLNHSORYAKIPLSPENKRG
IAFDHMKVCFQDPPTVPLKIPILKAFAGDIEEMVERRRKRRLRLVRAALLTQADSP
EDREPOQKRAHNDPERSPOKSESVANTIGARY"

BASE COUNT 8641 a 8030 c 8142 g 9780 t
ORIGIN

Query Match 1 9%: Score 46; DB 94; Length 34593;
Best Local Similarity 100.0%; Fred. No. 1.8e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 445 gaagcaatcaaatattctgtgcatgaaagaaactg 490
|||||
Db 10969 GAAGCAATCAAAATATTCTGTGTCATGGAAGCATGAACTCG 11014
|||||
RESULT 12
AC005545/C
LOCUS
DEFINITION
AC005545
VERSION
AC005545.1 GI:3478638
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 43514)
Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Adamson,A.W.,
Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stillwagen,S.,
Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garmes,J.,
Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Ayala,J.,
Liu,S., Attix,C., Andreise,T., Frankheim,M., Antico-Keller,G.,
Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
Krommiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
serine protease gene cluster
Unpublished
2 (bases 1 to 43514)
Lamerdin,J.E.
Direct Submission
Submitted (27-AUG-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
3 (bases 1 to 43514)
Lamerdin,J.E.
Direct Submission
Submitted (03-SEP-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
Map and sequence oriented from p telomere to centromere. Cosmid
R26634 overlaps cosmid R26660 (AC005328) to the left from bases 1
to 4,663 of this accession, and overlaps cosmid F6882 (AC005257) to
the right from bases 41,662 to 43,514. Additional chromosome 19 map
and sequence information may be obtained at:
http://www-bio.livl.gov/dbzr/genome/genome.html.
Location/Qualifiers
1. 43514
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R26634"
/chromosome="19"
/map="19p13.3 between CDC34 and DL9S342"
/cel_line="5H2-B"
/clone_lib="L119NC03 R chromosome 19-specific cosmid
library"
/note="cosmid library constructed at LLNL from flow-sorted
chromosomes from hybrid 5H2-B, which carries chromosome
19 as its only human chromosome."
46. 152
/note="BDS similarity to overlapping ESTs:
(46..152) AA431731 zw77g03.s1 Soares testis NHT Homo
sapiens cDNA clone 782260 3'; (391..340): 100%
identity.-(111..152) AA609572 af15c02.s1 Soares testis NHT
Homo sapiens cDNA clone 1031714 3'; (372..331): 100%
identity."
complement(273..608)
/note="BDS similarity to overlapping ESTs:
(608..273) T05239 EST03128 Homo sapiens cDNA clone
HEBCT75. Score: 621 Identity: 325/337 (96%)-(-718..308)
AA496009 zv72a03.s1 Soares total fetus Nd2H8 9w Homo
sapiens cDNA clone 759148 3'; Score: 755 Identity:
398/408 (97%)."
378..707
/note="BDS similarity to overlapping ESTs:

AA43171 zw77g03.s1 Soares testis NHT Homo sapiens CDNA clone 782260 3'; (339..11): 98% identity.-AA609572 af15c02.s1 Soares testis NHT Homo sapiens CDNA clone 1031714 3'; (330..5): 98% identity.-AA448181 zw83c05.s1 Soares testis NHT Homo sapiens CDNA clone 782792 3'; (336..4): 98% identity.

#

<378..468

/note="Hypothetical human protein with no significant similarity to known proteins"

/codon_start=2

/evidence=not_experimental

/product="R26660.2, partial CDS"

/protein_id="AAC34213.1"

/db_xref="GI:3478640"

/translation="LVSPALRCCEPHLANITLEDAAECLKH"

1771..3383

/note="IDS similarity to overlapping ESTs: (3391..3154) AA630776 ac14c10.s1 Stratagene Hela cell s3 937216 Homo sapiens CDNA clone 856434 3' similar to TR:G163328 G163328 LEUKEMIA VIRUS CELL RECEPTOR ; (262..1): 98% identity.- (1771..2449) AA167736 zq40004.s1 Stratagene hnt neuron (#937233) Homo sapiens CDNA clone 632167 3'; Score: 1207 Identity: 655/674 (97%)-(2109..2638) AA843415 ak07f04.s1 Soares parathyroid tumor NBHPA Homo sapiens CDNA clone IMAGE:1405279 3'; Score: 969 Identity: 516/547 (94%)-(2814..2510) AA64678 EST75336 pineal gland II Homo sapiens CDNA 5' end; Score: 582 Identity: 301/305 (98%)-(3200..2766) H73448 yu02f02.r1 Homo sapiens CDNA clone 232635 5'. Score: 717 Identity: 413/446 (92%)-(3358..2857) H28923 ym33c10.r1 Homo sapiens CDNA clone 49874 5'. Score: 756 Identity: 465/518 (89%)-(3383..2950) W69851 zd48c04.r1 Soares fetal heart NBH19W Homo sapiens CDNA clone 343878 5' similar to PIR:A45716 AA5716 Leukemia virus cell receptor ; Score: 744 Identity: 418/440 (95%)-Additional EST matches:

AA617872 AA716429 W63772 AA305631 AA321888 AA371945, R18826, T33092, T30887, F01584 and many others"

complement(2119..>39837)

/note="Human AP-3 complex delta subunit mRNA"

/product="Homo sapiens delta-adaptin mRNA"

complement(join(3296..3391,9810..9889,10209..10330,10986..11081,11259..11347,11830..12019,12408..12455,12802..12951,13983..14054,14459..14536,15248..15425,15871..15944,16342..16541,16661..16736,17330..17401,17728..17869,18345..18490,19724..19955,21985..22215,22286..22434,22857..23002,24481..24529,24953..25002,28275..28324,30213..30244,31625..31660,33594..33701,38134..38214,38850..38930,39742..>39837))

/note="vesicle coat component, similar to alpha and gamma adaptins"

/codon_start=1

/product="AP-3 complex delta subunit, partial CDS"

/protein_id="AAC34214.1"

/db_xref="GI:3522925"

/translation="AKTISOCIDEIKOELKODNIAVKAIVCKITLYIOMLGYDISMAA FNIEVMSAKFTFKRIGYLAASOSFHEGTDVIMLTNOIRKDISPSQYDTGVALG LSCVFPDLARDLANDIMTLMSHTKPYIRKRAKVIWVKVFLKDESRAPRPRKREK L AVILSSGMPNHSASIOLCVOKRLIEDSDOMKILGLGLAMSLIKTHPKSOSHK DILIOCLDDDESRILRALDILGVMSKKNMELVKKMLHVDKABEETPYRDELITK IDICSOINYOITTFEWTISILVELTFLBGRHGLLAQOMDLVAIRKARVAVSO MSALDSAHLLASTORNGICEVLYAAWTCGESEHLQEHHTLEMLRRVTLFG HIGAVYQNVVVKIYASILQOEAGEAGAOVQIOWDRIPOVGSADLEOVRASG ILQVKKHIQLOAKDVAIVAEVSALFAGELPVPAPKQKVPVEGSDILAMINPIS DSDEBDRPAVFEHEORRPKRHSSEADEELARREARROEDANNPEVTKSSPDS KRQDTGVSEHPIVYQIDLSVPLKVPGLPSMDQVYKLEERRRHOKLEKRRKRRRE KKKGRKRRHSLEPSEDEDIAPAOQVIVTEEMENALPSEDEKDRNDPRALDIDL DKPLADESKLPIQKHNTETSKPEKQVPAWEKSKPKKKKKKKKKERKKERKKER KKKKADLDFWLSTPPAPAPAPAPVSTDECEADKAEAGDEDAAGDODKSPR EKKKHKREKERTGKKKSKKOPGSEAGEPQONAPREEQPLPPESSYSLAENSY VKMTCDIRGSLQEDSOVTAIVLENNRSSILKGMELTLENNARMAPOSSVHDGV PVPRQLPRGVSNEMQVFTTQSIYMAKKIKGTLSTFIKNDGCAHFKELDFRLHSCS

CDS

YLTTPCYSDAPFAKLTESGDLSSMSIKVDIRMSFONLAKICFHNFHVSVERVDSCA SMYSRSIOGHVCLLVKGENSVSDGKSDSTLLNLEEMKATLAKC"

complement(join(3296..3391,9810..9889,10209..10330,10986..11081,11259..11347,11830..12019,12408..12455,12802..12951,15248..15425,15871..15944,16342..16541,16661..16736,17330..17401,17728..17869,18345..18490,19724..19955,21985..22215,22286..22434,22857..23002,24481..24529,24953..25002,28275..28324,30213..30286,30441..30580,31531..31660,33594..33701,38134..38214,38850..38930,39742..>39837))

/function="vesicle coat component"

/note="similar to alpha and gamma adaptins"

/function="AP-3 complex delta subunit"

/codon_start=1

/product="delta-adaptin, partial CDS"

/protein_id="AAC34212.1"

/db_xref="GI:3478639"

/translation="AKTISOCIDEIKOELKODNIAVKAIVCKITLYIOMLGYDISMAA FNIEVMSAKFTFKRIGYLAASOSFHEGTDVIMLTNOIRKDISPSQYDTGVALG LSCVFPDLARDLANDIMTLMSHTKPYIRKRAKVIWVKVFLKDESRAPRPRKREK L EDPGVOAAYVVICELARPNKPYILAPDKPIKMTSTNNVYIKIKIGCATLP LEPLGKRLLEPLNLHSTNSASLIECVNTYIAVILSSGMPNHSASIOLCVOKL RILIEDSDOMKILGLGLAMSLIKTHPKSVQSHKDLIOCLDDDESRILRALDILG MVSKKNMELVKKMLHVDKABEETPYRDELITKIDICSOINYOITTFEWTISILVE LTRLEGTRHGLLAQOMDLVAIRKARVAVSOMSALDSAHLLASTORNGICEV L YAAAMICGESEHPIVYQIDLSVPLKVPGLPSMDQVYKLEERRRHOKLEKRRKRRRE KKKGRKRRHSLEPSEDEDIAPAOQVIVTEEMENALPSEDEKDRNDPRALDIDL DKPLADESKLPIQKHNTETSKPEKQVPAWEKSKPKKKKKKKKKERKKERKKER KKKKADLDFWLSTPPAPAPAPVSTDECEADKAEAGDEDAAGDODKSPR EKKKHKREKERTGKKKSKKOPGSEAGEPQONAPREEQPLPPESSYSLAENSY VKMTCDIRGSLQEDSOVTAIVLENNRSSILKGMELTLENNARMAPOSSVHDGV PVPRQLPRGVSNEMQVFTTQSIYMAKKIKGTLSTFIKNDGCAHFKELDFRLHSCS

3435..3567

/rpt_family="AluSq"

3570..3866

/rpt_family="AluY"

3867..3906

/rpt_family="AluY"

3908..4090

/rpt_family="(TAA)n"

4101..4266

/rpt_family="AluSq"

complement(5793..6096)

/rpt_family="AluJo"

complement(6348..6855)

/rpt_family="MER9"

6938..7227

/rpt_family="AluSg"

7255..7402

/rpt_family="L1MB5"

7403..7689

/rpt_family="AluSx"

7693..8049

/rpt_family="L1MB5"

8105..8399

/rpt_family="AluY"

Query Match 0.94; Score 23; DB 86; Length 43514;
Best Local Similarity 100.0%; Pred. No. 0.43;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

07 653 atgaagaatgagccacacttca 675
|||||
Db 10491 ATGAATAATGAGCCACACTTCA 10469

RESULT 13
AF165142

LOCUS AF165142 105490 bp DNA PRI 24-AUG-1999
DEFINITION Homo sapiens chromosome 8 clone BAC 392C11 map 8p11-p12, complete
sequence.
ACCESSION AF165142
VERSION AF165142.1 GI:5499748
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutharia; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 105490)
Schudy,A., Blechschmidt,K., Schillhabel,M., Baumgart,C., Menzel,U.,
Weber,J., Schatevov,R. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1..105490
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8p11-p12"
/clone="BAC 392C11"
complement(84..140)
/note="GRAIL"
/evidence=not_experimental
repeat_region complement(333..592)
/rpt_family="AluSx"
716..761
/rpt_type=tandem
repeat_region complement(1299..1513)
/rpt_family="MIR"
3417..3519
/note="MZF"
/evidence=not_experimental
repeat_region complement(3617..4219)
/rpt_family="Charlie1a"
repeat_region complement(4209..4252)
/rpt_family="Charlie1a"
4302..4405
/note="GRAIL"
/evidence=not_experimental
exon complement(5085..5193)
/note="MZF"
/evidence=not_experimental
repeat_region complement(5315..5460)
/rpt_family="FLAM_C"
repeat_region complement(5347..5460)
/rpt_family="AluJo"
exon complement(5373..5537)
/note="MZF"
/evidence=not_experimental
exon 6488..6564
/note="GRAIL"
/evidence=not_experimental
repeat_region complement(6760..7029)
/rpt_family="AluY"
7180..7227
/rpt_type=tandem
repeat_region complement(7232..7489)
/rpt_family="AluSx"
exon complement(7489..7563)
/note="GRAIL"
/evidence=not_experimental
exon 8092..8142
/note="Initial_exon"
complement(8101..8149)
/note="GRAIL"
/evidence=not_experimental
exon 8306..8360
/note="GRAIL"
/evidence=not_experimental
repeat_region 9096..9528

repeat_region /rpt_family="Tigger2a"
9191..9229
/rpt_type=tandem
exon 9223..9282
/note="GRAIL"
/evidence=not_experimental
exon 9223..9282
/note="Terminal_exon"
repeat_region 10179..10635
/rpt_family="MER31B"
repeat_region 10824..10879
/rpt_type=tandem
repeat_region 10882..11145
/rpt_family="MER31-internal"
repeat_region 11138..11339
/rpt_family="MER31-internal"
repeat_region 11351..11994
/rpt_family="MER31-internal"
repeat_region 11563..11606
/rpt_type=tandem
repeat_region 12029..12223
/rpt_family="MER57-internal"
repeat_region 12122..12742
/rpt_family="MER31-internal"
repeat_region 12457..13937
/rpt_family="MER4-internal"
repeat_region 12637..13415
/rpt_family="MER65-internal"
repeat_region 12749..14506
/rpt_family="MER31-internal"
exon 13011..13027
/note="XPOUND"
/evidence=not_experimental
exon 14127..14173
/note="XPOUND"
/evidence=not_experimental
exon 14239..14331
/note="GRAIL"
/evidence=not_experimental
exon 14439..14470
/note="XPOUND"
/evidence=not_experimental
repeat_region 14513..14985
/rpt_family="MER31B"
repeat_region 14986..15205
/rpt_family="MER31-internal"
repeat_region 15583..15624
/rpt_type=tandem
repeat_region complement(15594..16226)
/rpt_family="AluYa8"
repeat_region complement(15778..16081)
/rpt_family="AluSc"
repeat_region complement(16414..16776)
/rpt_family="THE1B"
repeat_region complement(1761..18189)
/rpt_family="L1PA13"
repeat_region complement(18043..18271)
/rpt_family="L1"
repeat_region 18286..18554
/rpt_family="L1PA10"
exon 18291..18397
/note="MZF"
/evidence=not_experimental
exon 18291..18397
/note="GRAIL"
/evidence=not_experimental
repeat_region complement(18661..19071)
/rpt_family="MLT1F"
repeat_region complement(19781..20183)
/rpt_family="MER83"
exon 20013..20114
/note="GRAIL"
/evidence=not_experimental

```

repeat_region complement(21158, .21279)
                /rpt_family="MIR"
repeat_region 21802, .21906
                /rpt_family="MER5A"
repeat_region complement(22020, .22325)
                /rpt_family="Alub"
repeat_region 22911, .23014
                /rpt_family="MLT1J"
repeat_region 23103, .23196
                /rpt_family="L2a"
                complement(23581, .23725)
                /note="GRAIL"
exon 23583, .23663
        /evidence="not_experimental"
        /note="GRAIL"
exon 24026, .24185
        /evidence="not_experimental"
        /note="GRAIL"
exon complement(24093, .24257)
        /evidence="not_experimental"
        /note="Terminal_exon"
exon 24114, .24185
        /note="XPOUND"
exon 24201, .24344
        /evidence="not_experimental"
        /note="GRAIL"
repeat_region complement(24884, .24945)
                /rpt_family="L1MA6"
exon complement(25524, .25621)
        /note="XPOUND"
        /evidence="not_experimental"
repeat_region 25642, .25699
        /rpt_family="L2"
repeat_region 25837, .25924
        /rpt_type=tandem
repeat_region 26272, .27029
        /rpt_family="L1PA7"
exon 26733, .26881
        /note="GRAIL"
        /evidence="not_experimental"
exon 27656, .27783
        /note="GRAIL"
        /evidence="not_experimental"
exon complement(27765, .28027)
        /rpt_family="LTR37B"
repeat_region 28150, .28450
        /rpt_family="MER33"
repeat_region 28452, .29399
        /rpt_family="L1MA5"
repeat_region 28897, .29188
        /rpt_family="Alusx"
repeat_region 29402, .29593
        /rpt_family="Charlie5"
        /rpt_family="Charlie5"

Query Match      0.9%: Score 23; DB 89; Length 105490;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1428 aagaagtcagctaccagaataca 1450
    |||||
Db 39697 AAGAGTCAGTACCAGAAATCA 39719

RESULT 14
AC019331 190705 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 8 clone RP11-104D16, WORKING DRAFT
DEFINITION
SEQUENCE, 16 unordered pieces.
AC019331
AC019331.4 GI:85669783
VERSION
KEYWORDS
SOURCE human.
```

```

ORGANISM Homo sapiens
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS Mammalia: Eutheria: Primates: Catarrhini, Homiidae: Homo.
TITLE 1 (bases 1 to 190705)
JOURNAL Waterston, R.H.
REFERENCE The sequence of Homo sapiens clone
AUTHORS Unpublished
TITLE 2 (bases 1 to 190705)
JOURNAL Waterston, R.H.
COMMENT Direct Submission
Submitted (01-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 16, 2000 this sequence version replaced gi:7717162.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0104D16
----- Summary Statistics -----
Sequencing vector: M13; 49%
Sequencing vector: plasmid; 51%
Chemistry: Dye-primer ET; 49% of reads
Chemistry: Dye-terminator Big Dye; 51% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 182645 bases at least Q40
Consensus quality: 185230 bases at least Q30
Consensus quality: 186612 bases at least Q20
Insert size: 199000; agarose-fp
Quality coverage: 4.52 in Q20 bases; agarose-fp
Quality coverage: 4.79 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 2064: contig of 2064 bp in length
* 2065 2164: gap of unknown length
* 2185 3805: contig of 1641 bp in length
* 3806 3905: gap of unknown length
* 3906 8649: contig of 4744 bp in length
* 8650 8749: gap of unknown length
* 8750 13321: contig of 4572 bp in length
* 13322 13421: gap of unknown length
* 13422 19509: contig of 6088 bp in length
* 19510 19609: gap of unknown length
* 19610 24715: contig of 5106 bp in length
* 24716 24815: gap of unknown length
* 24815 30643: contig of 5828 bp in length
* 30644 30743: gap of unknown length
* 30744 40713: contig of 9970 bp in length
* 40714 40813: gap of unknown length
* 40814 48495: contig of 7682 bp in length
* 48496 48595: gap of unknown length
* 48596 57063: contig of 8468 bp in length
* 57064 57163: gap of unknown length
* 57164 67250: contig of 10087 bp in length
* 67251 67350: gap of unknown length
* 67351 82559: contig of 15209 bp in length
* 82560 82659: gap of unknown length
* 82660 102558: contig of 19899 bp in length
* 102559 102658: gap of unknown length
* 102659 124552: contig of 21894 bp in length
* 124553 124652: gap of unknown length
* 124653 152228: contig of 27576 bp in length
* 152229 152328: gap of unknown length
```

```

FEATURES      * 153329 190705: contig of 38377 bp in length.
source
1.190705
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="RP11-104D16"
1..2064
/note="assembly_name:Contig7"
2165..3805
/note="assembly_name:Contig8"
clone_end:SP6
vector_side:left"
3906..8649
/note="assembly_name:Contig9"
8750..13321
/note="assembly_name:Contig10"
13422..19509
/note="assembly_name:Contig11"
19610..24715
/note="assembly_name:Contig12"
24816..30643
/note="assembly_name:Contig13"
30744..40713
/note="assembly_name:Contig14"
40814..48495
/note="assembly_name:Contig15"
48596..57063
/note="assembly_name:Contig16"
57164..67250
/note="assembly_name:Contig17"
67351..82559
/note="assembly_name:Contig18"
82660..102558
/note="assembly_name:Contig19"
102659..124552
/note="assembly_name:Contig20"
124653..152228
/note="assembly_name:Contig21"
152329..190705
/note="assembly_name:Contig22"

BASE COUNT  55947 a 36657 c 35874 g 60716 t 1511 others
ORIGIN
Query Match      0.9%; Score 23; DB 65; Length 190705;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1428 aagaagctcagtaaccagaatca 1450
|||||
Db 161565 AAGAACTCAGTACCAGAAATCA 161587

RESULT 15
AL139399/c 64151 bp DNA PRI 21-DEC-2000
LOCUS Human DNA sequence from clone RP11-574A21 on chromosome
DEFINITION Xq21.33-22.3, complete sequence.
ACCESSION AL139399
VERSION AL139399.9 GI:11989998
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Howden,P.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Dec 22, 2000 this sequence version replaced gi:11878435.

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TrEMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/c-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX> RP11-574A21 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.choi.org/bacpac/home.htm> VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-574A21 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-274M8 is at 64052 in this sequence. The true right end of clone RP1-267M20 is at 100 in this sequence.

```

FEATURES      source
1..64151
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q21.33-22.3"
/clone="RP11-574A21"
/clone_lib="RPCI-11.2"

BASE COUNT  20073 a 11574 c 12196 g 20308 t
ORIGIN
Query Match      0.8%; Score 21; DB 90; Length 64151;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1351 ctgcagctcccaacttcag 1371
|||||
Db 24576 CTGCAGCTTCCCACTTCAG 24556

RESULT 16
AC008929 125403 bp DNA PRI 18-APR-2000
LOCUS Homo sapiens chromosome CTD-2291F22, complete sequence.
ACCESSION AC008929
VERSION AC008929.3 GI:6598820
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 125403)
REFERENCE DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-APR-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 125403)
REFERENCE DOE Joint Genome Center and Stanford Human Genome Center.
TITLE Direct Submission

```


JOURNAL Submitted (17-DEC-1999) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 4 (bases 1 to 125403)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (01-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 5 (bases 1 to 125403)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (02-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 6 (bases 1 to 125403)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Dec 17, 1999 this sequence version replaced gi:1615130.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu

Quality: Phrap Quality >=40 99.8% of Sequence;

Estimated Total Number of Errors is 0.3.

FEATURES
source 1. .125403
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="CTD-2291F22"
/clone_11b="Caltech human BAC library D"

BASE COUNT 39991 a 24654 c 23616 g 37142 t

ORIGIN

Query Match 0.8%; Score 21; DB 87; Length 125403;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 745 ttaagagaggaactcttg 765
|||||

Db 78704 TTAAGAGAGGAACTCTTG 78724

RESULT 17
AC027347/c AC027347 148807 bp DNA HTG 18-JUL-2000
LOCUS Homo sapiens chromosome 5 clone CTD-2562121, WORKING DRAFT
DEFINITION
SEQUENCE, 3 ordered pieces.
ACCESSION AC027347
VERSION AC027347.3 GI:9256715
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 148807)
AUTHORS DOE Joint Genome Institute.
TITLES
JOURNAL Unpublished
2 (bases 1 to 148807)
JOURNAL DOE Joint Genome Institute.
REFERENCE Direct Submission
JOURNAL Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7712058.

COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 814002
Center clone name: CITB-EL_2562121

Summary Statistics
Consensus quality: 147486 bases at least Q40
Consensus quality: 148415 bases at least Q20
Consensus quality: 148542 bases at least Q20
Estimated insert size: 234950; agarose-fp estimation
Estimated insert size: 148757; sum-of-ctgls estimation
Quality coverage: 7.48 in Q20 bases; agarose-fp estimation
Quality coverage: 11.81 in Q20 bases; sum-of-ctgls estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 85201: contig of 85201 bp in length
* 85202 85301: gap of unknown length
* 85302 103160: contig of 17859 bp in length
* 103161 103260: gap of unknown length
* 103261 148807: contig of 45547 bp in length.

FEATURES
source 1. .148807
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2562121"
/clone_11b="Caltech human BAC library D"

BASE COUNT 46028 a 28099 c 28931 g 45549 t

ORIGIN

Query Match 0.8%; Score 21; DB 70; Length 148807;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 745 ttaagagaggaactcttg 765
|||||

Db 5835 TTAAGAGAGGAACTCTTG 5815

RESULT 18
AL442223/c AL442223 162344 bp DNA HTG 24-JAN-2001
LOCUS Homo sapiens chromosome 1 clone RP11-426P2, *** SEQUENCING IN
DEFINITION
PROGRESS ***, 3 unordered pieces.
ACCESSION AL442223
VERSION AL442223.6 GI:11493327
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FUULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 162344)
AUTHORS Pavitt, R.
TITLES
JOURNAL Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Nov 30, 2000 this sequence version replaced gi:11493090.

COMMENT
-----Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----Project Information
Center project name: ba426P2
-----Summary Statistics
Assembly program: XGAP; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 160707 bases at least Q40

Consensus quality: 161514 bases at least Q30
 Consensus quality: 161929 bases at least Q20
 Insert size: 162144; sum-of-contigs
 Insert size: 166416; agarose-fp
 Quality coverage: 5.90x in Q20 bases; sum-of-contigs quality
 coverage: 6.66x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 33529: contig of 33529 bp in length
 * 33530 33629: gap of 100 bp
 * 33630 68844: contig of 35215 bp in length
 * 68845 68944: gap of 100 bp
 * 68945 162344: contig of 93400 bp in length.
 * location/Qualifiers
 * 1..162344
 * /organism="Homo sapiens"
 * /db_xref="taxon:9606"
 * /chromosome="1"
 * /clone="RP11-426P2"
 * /clone_1lb="RPC1-11.2"
 * 1..33529
 * /note="assembly-fragment:02253
 * fragment_chain:1"
 * 33630..68844
 * /note="assembly-fragment:01727
 * fragment_chain:1"
 * 68845..162344
 * /note="assembly-fragment:01094
 * clone_end:T7
 * vector_side:right"
 *
 * BASE COUNT 49214 a 32490 c 32393 g 48046 t 201 others
 * ORIGIN
 *
 * Query Match 0.8%; Score 21; DB 81; Length 162344;
 * Best Local Similarity 100.0%; Pred. No. 6.5;
 * Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 *
 * QY 338 tgcactgctatattgtggg 358
 * ||||||||||||||||||||
 * Db 31695 tgcactgctatattgtggg 31675
 *
 * RESULT 19
 * AP000719 196424 bp DNA PRI 22-FEB-2001
 * LOCUS Homo sapiens genomic DNA, chromosome 11q clone:RP11-757C15,
 * DEFINITION complete sequences.
 * ACCESSION AP000719
 * VERSION AP000719.4 GI:13094220
 * KEYWORDS HTG.
 * SOURCE Homo sapiens DNA, clone:RP11-757C15.
 * ORGANISM Homo sapiens
 * Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 * Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 *
 * REFERENCE
 * 1 (sites)
 * Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 * Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 * Homo sapiens genomic DNA
 * Published Only in Database (1999) in press
 * 2 (bases 1 to 196424)
 * Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 * Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 * Direct Submission
 * Submitted (11-NOV-1999) Masahira Hattori, The Institute of Physical
 * and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 On Feb 21, 2001 this sequence version replaced gi:5927288.
 location/Qualifiers
 1..196424
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q"
 /clone="RP11-757C15"
 *
 * BASE COUNT 54557 a 42826 c 43446 g 55595 t
 * ORIGIN
 *
 * Query Match 0.8%; Score 21; DB 91; Length 196424;
 * Best Local Similarity 100.0%; Pred. No. 6.5;
 * Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 *
 * QY 976 ttccagaggtaccagaagaag 996
 * ||||||||||||||||||||
 * Db 160132 ttccagaggtaccagaagaag 160152
 *
 * RESULT 20
 * AP002490/c 196613 bp DNA HTG 10-NOV-2000
 * LOCUS Homo sapiens chromosome 11 clone CMB9-17C1 map 11q12, WORKING DRAFT
 * DEFINITION SEQUENCE, 25 unordered pieces.
 * ACCESSION AP002490
 * VERSION AP002490.2 GI:11138084
 * KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
 * SOURCE Homo sapiens DNA, clone:CMB9-17C1.
 * ORGANISM Homo sapiens
 * Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 * Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 *
 * REFERENCE
 * 1 (bases 1 to 196613)
 * Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 * Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 * Homo sapiens 196,613 genomic DNA of 11q12
 * Published Only in Database (2000) in press
 * 2 (bases 1 to 196613)
 * Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 * Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 * Direct Submission
 * Submitted (09-JUN-2000) Masahira Hattori, The Institute of Physical
 * and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 * 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 * (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 * Tel:81-45-503-9111, Fax:81-45-503-9170)
 * On Nov 11, 2000 this sequence version replaced gi:5547575.
 *
 * COMMENT
 * ----- Genome Center
 * Center: RIKEN Genomic Sciences Center(GSC)
 * Center code: RIKEN
 * Web site: http://hgp.gsc.riken.go.jp/
 * Contact: hattori@gsc.riken.go.jp
 * ----- Project Information
 * Center project name: HumDrafft11
 * Center clone name: CMB9-17C1
 * ----- Summary Statistics
 * Sequencing vector: PCR products; 100% of reads
 * Chemistry: Dye-terminator ET-amersham; 100% of reads
 * Assembly program: Phrap; version 0.990329
 * Consensus quality: 190034 bases at least Q40
 * Consensus quality: 192516 bases at least Q40
 * Consensus quality: 193548 bases at least Q20
 * Insert size: 194213; sum-of-contigs
 * quality coverage: 8.73x in Q20 bases; sum-of-contigs
 * -----
 * NOTE: This is a 'working draft' sequence. It currently consists of
 * 25 contigs. The true order of the pieces is not known and their
 * order in this sequence record is arbitrary. Gaps between the
 * contigs are represented as runs N, but the exact sizes of the gaps

are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1 27851 contig of 27851 bp in length
27952 58349 contig of 30388 bp in length
58450 76029 contig of 17580 bp in length
76130 92048 contig of 15919 bp in length
92149 103650 contig of 11502 bp in length
103751 114318 contig of 10568 bp in length
114419 123627 contig of 9209 bp in length
123728 131291 contig of 7564 bp in length
131392 139816 contig of 8425 bp in length
139917 146953 contig of 7037 bp in length
147054 153039 contig of 5986 bp in length
153140 158083 contig of 4944 bp in length
158184 163158 contig of 4975 bp in length
163159 163258 gap of 100 bp in length
163259 169022 contig of 5764 bp in length
169023 169122 gap of 100 bp in length
169123 171598 contig of 2476 bp in length
171599 171698 gap of 100 bp in length
171699 175793 contig of 4095 bp in length
175794 175883 gap of 100 bp in length
175894 178616 contig of 2723 bp in length
178617 178716 gap of 100 bp in length
178717 181798 contig of 3082 bp in length
181799 181898 gap of 100 bp in length
181899 184417 contig of 2519 bp in length

1 27851: contig of 27851 bp in length
27852 27951: gap of 100 bp
27952 58349: contig of 30388 bp in length
58350 58449: gap of 100 bp
58450 76029: contig of 17580 bp in length
76030 76129: gap of 100 bp
76130 92048: contig of 15919 bp in length
92049 92148: gap of 100 bp
92149 103650: contig of 11502 bp in length
103651 103750: gap of 100 bp
103751 114318: contig of 10568 bp in length
114319 114418: gap of 100 bp
114419 123627: contig of 9209 bp in length
123628 123727: gap of 100 bp
123728 131291: contig of 7564 bp in length
131292 131391: gap of 100 bp
131392 139816: contig of 8425 bp in length
139817 139916: gap of 100 bp
139917 146953: contig of 7037 bp in length
146954 147053: gap of 100 bp
147054 153039: contig of 5986 bp in length
153040 153139: gap of 100 bp
153140 158083: contig of 4944 bp in length
158084 158183: gap of 100 bp
158184 163158: contig of 4975 bp in length
163159 163258: gap of 100 bp
163259 169022: contig of 5764 bp in length
169023 169122: gap of 100 bp
169123 171598: contig of 2476 bp in length
171599 171698: gap of 100 bp
171699 175793: contig of 4095 bp in length
175794 175883: gap of 100 bp
175894 178616: contig of 2723 bp in length
178617 178716: gap of 100 bp
178717 181798: contig of 3082 bp in length
181799 181898: gap of 100 bp
181899 184417: contig of 2519 bp in length

NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
```

```
* 184418 184517: gap of 100 bp
* 184518 187490: contig of 2973 bp in length
* 187491 187590: gap of 100 bp
* 187591 189266: contig of 1678 bp in length
* 189269 189368: gap of 100 bp
* 189369 191568: contig of 2200 bp in length
* 191569 191668: gap of 100 bp
* 191669 193333: contig of 1665 bp in length
* 193334 193433: gap of 100 bp
* 193434 195156: contig of 1723 bp in length
* 195157 195256: gap of 100 bp
* 195257 196613: contig of 1357 bp in length.
location/Qualifiers
```

FEATURES

source

```
1. 196613
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="11"
  /map="11q12"
  /clone="CMB9-17C1"
1. 27851
  /note="assembly_fragment"
27952..58349
  /note="assembly_fragment"
58450..76029
  /note="assembly_fragment"
76130..92048
  /note="assembly_fragment"
92149..103650
  /note="assembly_fragment"
103751..114318
  /note="assembly_fragment"
114419..123627
  /note="assembly_fragment"
123728..131291
  /note="assembly_fragment"
131392..139816
  /note="assembly_fragment"
139917..146953
  /note="assembly_fragment"
147054..153039
  /note="assembly_fragment"
153140..158083
  /note="assembly_fragment"
158184..163158
  /note="assembly_fragment"
163259..169022
  /note="assembly_fragment"
169123..171598
  /note="assembly_fragment"
171699..175793
  /note="assembly_fragment"
175894..178616
  /note="assembly_fragment"
178717..181798
  /note="assembly_fragment"
181899..184417
  /note="assembly_fragment"
184518..187490
  /note="assembly_fragment"
187591..189268
  /note="assembly_fragment"
189369..191568
  /note="assembly_fragment"
191669..193333
  /note="assembly_fragment"
193434..195156
  /note="assembly_fragment"
195257..196613
  /note="assembly_fragment"
```

```
BASE COUNT 55160 a 43875 c 44156 g 51021 t 2401 others
ORIGIN
```



```

misc_feature 216239..220146
            /note="assembly_name:Contig9"
BASE COUNT 66033 a 41274 c 41775 g 69151 t 1913 others
ORIGIN

Query Match      0.8%; Score 21; DB 71; Length 220146;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 ttaagagagaacctcttg 765
      |||
Db 51116 TTAAGAGAGCAACTCTTGT 51136

RESULT 22
AB046207 1176 bp mRNA VRT 11-APR-2001
LOCUS Anguilla japonica mRNA for 28kDa-1d apolipoprotein, complete cds.
DEFINITION AB046207
ACCESSION AB046207
VERSION AB046207.1 GI:13591607
KEYWORDS
SOURCE Anguilla japonica liver cDNA to mRNA.
ORGANISM Anguilla japonica
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Anguilliformes;
Anguillidae; Anguillidae; Anguilla.
REFERENCE 1 (sites)
AUTHORS Kondo,H., Kawazoe,I., Nakaya,M., Kikuchi,K., Aida,K. and Watabe,S.
TITLE The novel sequences of major plasma apolipoproteins in the eel
Anguilla japonica(1)
JOURNAL Biochim. Biophys. Acta 1531 (1-2), 132-142 (2001)
MEDLINE 21175592
REFERENCE 2 (bases 1 to 1176)
AUTHORS Watabe,S., Kondo,H., Kawazoe,I., Nakaya,M., Kikuchi,K. and Aida,K.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2000) Shugo Watabe, The University of Tokyo,
Graduate School of Agricultural and Life Sciences; Yayoi 1-1-1,
Bunkyo, Tokyo 113-8657, Japan
(E-mail:watabe@mail.ecc.u-tokyo.ac.jp, Tel:81-3-5841-7520,
Fax:81-3-5841-8166)
FEATURES
source location/Qualifiers
1..1176
/organism="Anguilla japonica"
/db_xref="taxon:7937"
/lssue_type="liver"
18..797
/codon_start=1
/product="28kDa-1d apolipoprotein"
/protein_id="BAB0964.1"
/db_xref="GI:13591608"
/translation="MKPIVIVIALVSFTGCANLRSNPPQLELVDAEMDFVKAS
QTAGALKRTGESELAQVNAKIKESVEVQYRIQEOGVIVISDELHKLSHAG
LSESLQDINEVROLEPLAEKLSANIQOQKRPVLDPYTESLDLRALERLRMH
RLRMSVEQLSLDLEHLQDQSGTEELRKRVESYOEYREFLPLENIIGNILNK
LFMFRVYVYVDSLRKLDPIYTGLESQLTALWESSANS"
BASE COUNT 333 a 262 c 286 g 295 t
ORIGIN

Query Match      0.8%; Score 20; DB 8; Length 1176;
Best Local Similarity 100.0%; Pred. NO. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 agagcgtgcagaggtacag 1392
      |||
Db 601 AGAGCGTCAGAGTACAGG 620

RESULT 23
ATU23796 1573 bp mRNA PLN 21-JUL-1995
LOCUS Arabidopsis thaliana IAA-amino acid hydrolase homolog ILL2 (ILL2)
DEFINITION
```

```

mRNA, complete cds.
U23796
U23796.1 GI:902790
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1573)
AUTHORS Bartel,B. and Fink,G.R.
TITLE IRL1, an amidohydrolase that releases active indole-3-acetic acid
from conjugates
JOURNAL Science 268 (5218), 1745-1748 (1995)
MEDLINE 95312867
REFERENCE 2 (bases 1 to 1573)
AUTHORS Bartel,B. and Fink,G.R.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-1995) Bonnie Bartel, Whitehead Institute, 9
Cambridge Center, Cambridge, MA 02142, USA
On Jul 21, 1995 this sequence version replaced gi:887788.
FEATURES
source location/Qualifiers
1..1573
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/chromosome="5"
1..1320
/gene="ILL2"
/note="ILL2-like gene; IAA-amino acid hydrolase homolog"
/codon_start=1
/product="ILL2"
/protein_id="AAC49016.1"
/db_xref="GI:902791"
/translation="MALNKLISLTFQLLFLSVSSESPWIAEDTSIQTKLEFAPS
PEVDMVVKIRKIRHNPELGEYELFETSKIRSLRSLGKIRYPAIVIGVIGTG
EPPVVALRADMDALPIQEGVEHEKSKIPGMHACGHDGVTMLGAAKIIEHRHL
OGTVVLIPOAEESLGSKKMRGALKNVEATPGIHSARIPRGKASRSFLAGA
GVEFNAVITGKGNAIPIHHTIDPVYVAASSVLSLQQLVSRETDPLDSKYTVSKVNG
NAFNVIPDSITIGTLRAFTGFTQLQQRVVEVITKQAVHRCNANVLTDPGRPMPP
TVNNKDLVKQFKVRVDLQGEAFVEAPVWGSDFEFATETIPGHSLLGMODETNG
YASHSPLRYRNEDVLPYGAAIIRHSMVQYLKERSKGSVSGFHEEL"
75..1320
/gene="ILL2"
BASE COUNT 452 a 332 c 363 g 426 t
ORIGIN

Query Match      0.8%; Score 20; DB 14; Length 1573;
Best Local Similarity 100.0%; Pred. NO. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1703 ctttggaagcgcctcac 1722
      |||
Db 609 CTTTGGAAAGCCGCTTAC 628

RESULT 24
AF047031 4918 bp DNA PLN 03-MAY-1999
LOCUS Arabidopsis thaliana IAA-amino acid hydrolase (ILL1 and ILL2)
DEFINITION genes, complete cds.
ACCESSION AF047031
VERSION AF047031.1 GI:2921828
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 4918)
AUTHORS Bartel,B. and Fink,G.R.
TITLE IRL1, an amidohydrolase that releases active indole-3-acetic acid
```

from conjugates
JOURNAL Science 268 (5218), 1745-1748 (1995)
MEDLINE 95312867
2 (bases 1 to 4918)
AUTHORS Davies,R.T., Goetz,D.H., Lasswell,J., Anderson,M.N. and Bartel,B.
TITLE IAR3 encodes an auxin conjugate hydrolase from Arabidopsis
JOURNAL MEDLINE 99172052
3 (bases 1 to 4918)
AUTHORS Bartel,B.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1998) Biochemistry and Cell Biology, Rice
University, 6100 S. Main St., Houston, TX 77005, USA
Location/Qualifiers
source 1..4918
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/map="between nga129 and LFY3"
join(<143..496,605..730,805..1119,1201..1323,1408..>1806)
/gene="ILL1"
/product="IAA-amino acid hydrolase"
<143..>1806
/gene="ILL1"
/note="ILR1-like gene 1"
join(143..496,605..730,805..1119,1201..1323,1408..1806)
/gene="ILL1"
/function="auxin conjugate hydrolase"
/codon_start=1
/product="IAA-amino acid hydrolase"
/protein_id="AAC04865.1"
/db_xref="GI:2921839"
/translation="MALNMFLEFOLLILLRVSSSEPMIIVAGDVRIPINLEAKSP
EYDMSVRIRRIHSENELGEEPEENSKPIRSEDLIGVRYEPVINTGIIYIGGE
PPYVALRADMDALPIOEAVMEHREKSKPKMAACGDIQVAMILGAKILQORHQLQ
GTVVLFQPAEEGLSGAKMREGALKNEVALFGIHLSPRTPGKASLAGSFMAGAG
AEFAVITGKGHNAIPQHTIDPVVAASSIVLSIQHLVSRSTPDSKVTYTVKNGN
AFNVDITIGTIGTLRAFTGTOLOORKEVITTKQAVHRCNASVNLPPNGEMPP
VNMMDLYKKFKVVDLGOEAFVEAVPEMSDEFSYFAETIPGHFSLIGMDDETCY
ASSSPHYRINEDVLPYGAIIHATMAVOYLKAKSGSYSGFDEL"
join(<2596..2952,3121..3246,3323..3637,3718..3840,
3931..>4329)
/gene="ILL2"
/product="IAA-amino acid hydrolase"
<2596..>4329
/note="ILR1-like gene 2"
join(2596..2952,3121..3246,3323..3637,3718..3840,
3931..4329)
/gene="ILL2"
/function="auxin conjugate hydrolase"
/codon_start=1
/product="IAA-amino acid hydrolase"
/protein_id="AAC04866.1"
/db_xref="GI:2921830"
/translation="MALNKLILSTFOLLFLLSVSSSEPMIAEDTSQIOTKILFEAKS
PEYFDMVIRIRRIHSENELGEEPEENSKPIRSEDLIGVRYPAVIAITLHHRHNL
EPYVALRADMDALPIOEAVMEHREKSKPKMAACGDIQVAMILGAKILQORHQLQ
GTVVLFQPAEEGLSGAKMREGALKNEVALFGIHLSPRTPGKASLAGSFMAGAG
AEFAVITGKGHNAIPQHTIDPVVAASSIVLSIQHLVSRSTPDSKVTYTVKNGN
AFNVDITIGTIGTLRAFTGTOLOORKEVITTKQAVHRCNASVNLPPNGEMPP
VNMMDLYKKFKVVDLGOEAFVEAVPEMSDEFSYFAETIPGHFSLIGMDDETCY
ASSSPHYRINEDVLPYGAIIHATMAVOYLKAKSGSYSGFDEL"
variation 3082..3083
/gene="ILL2"
/note="6 base insertion in ecotype Landsberg erecta"
/replace="tgattt"
BASE COUNT 1422 a 991 c 997 g 1508 t
ORIGIN
Query Match 0.8%; Score 20; DB 12; Length 4918;

Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1703 cttgggaagacgccttcac 1722
|||||
Db 3448 CTTTGGGAAGCCGCTTCAC 3467
RESULT 25
AB014516
LOCUS AB014516 6718 bp mRNA PRI 06-FEB-1999
DEFINITION Homo sapiens mRNA for KIAA0616 protein, partial cds.
ACCESSION AB014516
VERSION AB014516.1 GI:3327045
KEYWORDS
SOURCE
Homo sapiens adult male brain cDNA to mRNA, clone_1lb:pbuescriptII
SK plus clone:H603623.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 6718)
AUTHORS Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-1998) to the DDBJ/EMBL/GenBank databases. Osamu
Sumata, Kazusa DNA Research Institute, Laboratory of DNA Technology;
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
2 (sites)
REFERENCE Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. X.
MEDLINE The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro
JOURNAL DNA Res. 5 (3), 169-176 (1998)
98403880
Location/Qualifiers
source 1..6718
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="H603623"
/clone_1lb="pbuescriptII SK plus"
/dev_stage="adult"
/sex="male"
/issue_type="brain"
1..1906
/gene="KIAA0616"
<1..1906
/gene="KIAA0616"
/codon_start=2
/product="KIAA0616 protein"
/protein_id="BAA31591.1"
/db_xref="GI:3327046"
/translation="NNOKAEEFAPEYMKDLSIRARLQLOKSOYLQGPSRGQY
YCGSLPNVNIQISGTHMDLFPQSPGEGELAAAPVSLTPQSSGIDTSKTHHRLVD
RVYRERGRVSGPHRPLSVDKHGKQADSCPYTWTLSIPADTSMKRTNSDLHGSTM
PTJOESESFGSQDYOQKRVLLITVPGMETTISEADKNLSKQAMDTKKGSPKCEV
GINTPEPSADQNTTALIPATNHTGSLDPLDNIHPPSLPLDPEETFPALSSS
STGNLAANTLHIGIGAGGGMSTPGSSPHRAGVPSLSSTEARQOASPTLPSLP
ITONAYAMDLSLEQOLPYAFYTOASQCPPOPPPPPPASQOAPPPPPPOADAVRL
PPGGLPSPASLIRGPPPLAVTPSSLPQSPENKPGQPSNGITIASAPALQOVRTS
AASPANQSTPSVSNQGFSGSPQHTSTLGSVFDVAYEQQMAARQAAALSHOLEOF
NMNENAISSSSISPSSTLNTYQOAMMGITLGSGLSPSQOQGVASHGIPNITLVT
GSPSLSKELTSLSLAGVDVSFGPSQGPPLDELKIDPLTLGLHMLNDPVDVLADPA
TDDTRMDRL"

BASE COUNT 1208 a 2237 c 1930 g 1343 t
ORIGIN
Query Match 0.8%; Score 20; DB 85; Length 6718;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1601 99gtccttgagcaccctgact 1620
|||||
Db 3731 GGGTCTTGCGCACCTGGCT 3750

RESULT 26
SC7B7/c
LOCUS Streptomyces coelicolor cosmid 7B7.
ACCESSION AL009199.1 GI:2661675
VERSION 1
KEYWORDS ABC-transport system; hydrolase; periplasmic solute-binding protein; transcriptional repressor protein; transmembrane transport protein.

SOURCE
ORGANISM Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 13800)
Murphy L. and Harris D.
Unpublished
2 (bases 1 to 13800)
Parkhill, J., Barrell, B.G. and Rands, M.A.
Direct Submission
Submitted (25-NOV-1997) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
3 (bases 1 to 13800)
Redenbach, M., Kleser, H.M., Denapate, D., Elchner, A., Cullum, J., Kinashi, H. and Hopwood, D.A.
A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 7B7 is in the AseI-B fragment, between 7H1 and 1C3.
Location/Qualifiers
1. 13800
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"

gene
CDS
misc_feature
gene
CDS
misc_feature
RBS
RBS
RBS
gene
CDS
misc_feature
/clone="cosmid 7B7"
complement(1..1878)
/gene="SC7B7.01c"
complement(<1..1878)
/gene="SC7B7.01c"
/note="SC7B7.01c, unknown ATP/GTP binding protein, partial ORF, len: >626 aa; contains possible N-terminal signal sequence and PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=1
/product="hypothetical ATP/GTP binding protein"
/protein_id="CAA15781.1"
/db_xref="GI:2661675"
/db_xref="SPTRMBL:050498"
/translation="MQYVRHRCWGPAAVMTTAAALAVSGCAAYALFGAPAKADREFA SFEADDPVDMTDTVDTDGGRKASGVGGYSTGIPGVTENVTGVRSAEMTGGREVKENLVNDEPATKMLTFAPTGVEFLDAPTKIAKALTSANDHDERDPLDLKGS ADGEMQTLDTSGEHDFGRFQRTYELAEFAERHRLRTTKNGAGDALOLADVOL ATSDTEVPTPEDMLSFYDEGPGSPTRKANAAGFGKALRAGHRTADGRAYSINKYF DVNKKVRSKELSTKTFPSMADGDLIDATNVSVDLAFDTGTYLSDLRATDQGEPLT PRGGDAKILYVQMNSVKSIGSVAAGTVDIRLLAYDSRPGAPKRGMLDVAIER AAPRRKHLSDYALTRGTNSGSGFRGNFPATAVPHGNEMPTPTNAGSLSWIE YASGNNDNLPTVOAFSASHPSPMGDRQTFQVMPSPAASGTPGCDARLARHEN ETARPYVGVFRENGKAKEMPTPDHAMMPTYPGDASVIFDWNDAQGLTLDKENG TFGYSVDRSGSGAATRLRYVGFSDRSVIGDSSGVKGHHRDPAGDHSTMTATRS LI"
complement(787..810)
/gene="SC7B7.01c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(1890..1894)
/note="possible RBS upstream of SC7B7.01c"
2428..2431
/note="possible RBS upstream of SC7B7.02"
2440..3849
/gene="SC7B7.02"
2440..3849
/gene="SC7B7.02"
/note="SC7B7.02, unknown, len: 469; has similarity to signal sequence of Mycobacterium leprae hypothetical protein TR:E34742 (EMBL:Z98604) MCB2052.27 (445 aa), fasta scores: opt: 75 z-score: 199.2 E(): 0.00062, 23.3% identity in 335 aa overlap. Contains N-terminal signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
/codon_start=1
/transl_table=1
/product="hypothetical protein SC7B7.02"
/protein_id="CAA15782.1"
/db_xref="GI:2661677"
/db_xref="SPTRMBL:050499"
/translation="MTIRAGSLDRRLTLRGAIVTANGSFAVACSSPSSDEKESDGP KGEKSANPFGAANSSTVEAALFDGGYGVYDVANQVLSGVKGLVQYKPVVDIAP OLOPRFVGNPDLIDNSGEODIGFLILDLELDLFEASTYEGKRIIVPGVK DGPFRKFPVALNVMTVYGVWSKTFEENGWTPRTMPEDALDGGEARKKGYLFV HGRKAATYRFTALIDLSAIKGCGDVRALLENLEKGCMSHRAVGVITVMEVMYQKMF VPGSGGTQFORQAATNSNDQKALITPISGVITENMKRAITRADFOMTIPSKTLTKRA LLYEALAAAGSEPTIVPKQGNPKAGKGVKRAMISEKAAANFSKTLAPTVKGTVA LDGYSTALVSQTKMLEAGATNIFNMFVEYGLNTDOLVPMNSFIAGDLGKGLTSAL OKISDKVRSDSVKRVYS"
2497..2529
/gene="SC7B7.02"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
3857..4918
/gene="SC7B7.03"
3857..4918
/gene="SC7B7.03"
/note="SC7B7.03, probable integral membrane transport protein, len: 353 aa; similar to many members of the MALFG subfamily of the binding-protein-dependent transport system membrane proteins eg. MSMP_STRMU 000750 multiple sugar-binding transport system permease protein (290 aa),

fasta scores; opt: 393 z-score: 636.8 E(): 2.6e-28, 32.5% identity in 280 aa overlap. Contains PS00402 binding-protein-dependent transport systems inner membrane comp signature"
/codon_start=1
/transl_table=11
/product="probable transmembrane transport protein"
/protein_id="CA15783.1"
/db_xref="GI:2661678"
/db_xref="SPTREMBL:O50500"
/translation="MKDPTPAETASRREPARGRRPKITPDVTEFLAFLGV LAIFVFLIPPGQALFPGMTDMRGSPRYNFGDPTFKMODIFFLAKNALVLA AVEPLVTLTLAGVAVALITLGGPSKGPVKGIGASRYRITSEPPYVPAIIVGLMAQ MVDPNAGLNGVLTGLDQPTFAMVIGKAAMPVMMVIGLVGFYAVFLAIK GVPGLYEAKIDGAGERRITITLPAIRDSVQFAYIGIALADAFYVOAMPNG GPDNSTLTISORLFINVAFKQOGYATANGVLAATVFAALVFLVNRITGGSEES KRKPGSRARRAAKNGAR"
4505..4591
/gene="SC7b7.03"
/note="PS00402 Binding-protein-dependent transport systems inner membrane component signature"
4904..4908
/gene="SC7b7.03"
/note="possible RBS upstream of SC7b7.04"
4915..5823
/gene="SC7b7.04"
4915..5823
/gene="SC7b7.04"
/note="SC7b7.04, probable integral membrane transport protein, len: 301 aa; similar to many members of the MALFG subfamily of the binding-protein-dependant transport systems membrane proteins eg. TR:Q0611 (EMBL:X66092) putative transport system permease protein from Clostridium perfringens (275 aa); fasta scores: opt: 590 z-score: 796.5 E(): 0.33.6% identity in 265 aa overlap. Contains PS00402 Binding-protein-dependent transport systems inner membrane comp signature"
/codon_start=1
/transl_table=11
/product="probable transmembrane transport protein"
/protein_id="CA15784.1"
/db_xref="GI:2661679"
/db_xref="SPTREMBL:O50501"
/translation="MSVIAADRRPASDRKPVSDRLARAASSDRRFAISHALLITM SVIIVPMVMVIMSSFKSGELISSPFLPDHMRFPENYANMTDANIGYFLNSYIV VVALITVMLGAMCAVYLARFEPGRRLIYYVLAGLFPVPLAIYPLFEOINSGIL NTRRGILITVYAFALPFTWFTIXSPFRSLPHYVYKALIDGAGDMRAFQVMIPNARP GAAVAITENFELGIMQFLIPVALNTDQDKWVLVQGAAYASSQVYDIDYGALFAITV TVVPLVLYCVQRIAGSVSGTFR"
5470..5556
/gene="SC7b7.04"
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp signature"
5851..5900
/note="hairpin loop with 23bp stem"
6044..6047
/note="possible RBS upstream of SC7b7.05"
6058..7269
/gene="SC7b7.05"
6058..7269
/gene="SC7b7.05"
/gene="SC7b7.05"

Query Match 0.8%; Score 20; DB 3; Length 13800;

Best local similarity 100.0%; Pred No. 23;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 77 ccgcccccgagcgagcgccg 96
|||||

Db 3925 CCGCCGCCGCGGAGCGCGC 3906

RESULT 27
AC003107

LOCUS AC003107 46275 bp DNA PRI 18-NOV-1997
DEFINITION Human DNA from chromosome 19-specific cosmid R30064 containing the COMP gene, genomic sequence, complete sequence.
ACCESSION AC003107
VERSION AC003107.1 GI:2623749
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 46275)
Lamerdin,J.E., McCreedy,P.M., Adamson,A.W., Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stillwagen,S., Gaines,D., Dangnanan,L., Christensen,M., Bruce,R., Quan,G., Montgomery,M., Ow,D., Kobayashi,A., Olsen,A.O. and Cariano,A.V.
Sequence analysis of an ~1 Mb region containing the MFE2B gene in 19p12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 46275)
Lamerdin,J.E.
AUTHORS Direct Submission
TITLE Submitted (18-NOV-1997) Human Genome Center, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
JOURNAL Cosmid R30064 overlaps cosmid F19807 to the left and cosmid R32469 to the right.
COMMENT
Map and sequence oriented from telomere to centromere.
FEATURES
source
Location/Qualifiers
1..46275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R30064"
/chromosome="19"
/map="19p12 between UBA52 and D19S451"
/cell_line="5H12-B"
/clone_11b="L119NC03 R chromosome 19 cosmid library"
/note="L119NC03 cosmid library constructed at LNL from flow-sorted chromosomes from hybrid 5H12-B, which carries chromosome 19 as its only human chromosome."
817..1037
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 93.000"
complement(3847..3899)
/rpt_family="MER7"
complement(3910..4198)
/rpt_family="Alu"
complement(4214..4315)
/rpt_family="Alu"
6213..6337
/rpt_family="MER7"
/note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: excellent, score: 93.000"
6488..6785
/rpt_family="Alu"
complement(6934..7242)
/rpt_family="Alu"
complement(7269..7545)
/rpt_family="Alu"
complement(7989..8304)
/rpt_family="Alu"
9294..9409
/note="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: good, score: 65.000"
9543..9602
/note="DGS similarity to AA378771 EST91505 synovial sarcoma Homo sapiens cDNA 5' end (1..60); 100% identity."
9676..10030
/note="DGS similarity to AA209234 zq85f01.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone 648409 5' (1..351); Score: 660 Identity: 346/351 (98%)."
complement(10540..11045)
/note="DGS similarity to AA211938 zq85f01.s1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone 648409 3' Score: 951 Identity: 497/501 (99%)."

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc_feature 11916..12187
/note="DDS similarity to T82171 yd95g06.r1 Homo sapiens cDNA clone 116025 5' similar to contains TARI repetitive element.Score: 4/8 Identity: 262/274 (95%)."

Other overlapping matches:
(11993..12136) predicted exon, program: grail2exons_human_1.3, frame: 2, quality: excellent, score: 81.000
(12176..11917) DDS similarity to T95404 ye43901.r1 Homo sapiens cDNA clone 120528 5'. Score: 471 Identity: 258/263 (98%).~(12713..11917) DDS similarity to T84000 yd66d10.r1 Homo sapiens cDNA clone 113203 5'. Score: 400 Identity: 257/275 (93%).~(11965..12181) DDS similarity to T95320 ye43901.s1 Homo sapiens cDNA clone 120528 3' similar to gb:M64241 QM PROTEIN (HUMAN).Score: 334 Identity: 203/218 (93%)."
12251..12355
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: good, score: 52.000-DDS similarity to AA378771 EST91505 Synovial sarcoma Homo sapiens cDNA 5' end (61..165). 96% identity."
misc_feature 13676..13705
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: good, score: 54.000"
14355..14457
/note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: excellent, score: 77.000"
15674..15795
/note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: good, score: 65.000"
15709..15764
/note="DSS similarity to AA378771 EST91505 Synovial sarcoma Homo sapiens cDNA 5' end (166..221): 99% identity."
misc_feature 16450..16630
/note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: excellent, score: 100.000"
17229..17316
/note="DSS similarity to T95918 ye42a04.r1 Homo sapiens cDNA clone 120366 5'. Score: 163 Identity: 86/87 (98%)."
18023..18187
/note="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: excellent, score: 100.000"
18533..18757
/note="DSS similarity to AA252605 zsl14b02.s1 NCI CGAP_CG81 Homo sapiens cDNA clone 685131 3' end. Score: 424 Identity: 225/225 (100%)."
20682..22267
/note="DSS similarity to overlapping ESTs:
(20682..21075) AA350958 EST58512 Infant brain Homo sapiens cDNA 5' end. Score: 776 Identity: 391/394 (99%).
(20863..21259) R87396 ym88h07.r1 Homo sapiens cDNA clone 166045 5'. Score: 748 Identity: 391/401 (97%).
(20874..21249) AA349930 EST57030 Infant brain Homo sapiens cDNA 5' end. Score: 748 Identity: 375/376 (99%).
(21125..21574) R19558 y926h07.r1 Homo sapiens cDNA clone 33726 5'. Score: 811 Identity: 440/457 (96%).
(21589..22030) H30788 y079f06.r1 Homo sapiens cDNA clone 184163 5' similar to gb:M33326 NONSPCIFIC CROSS-REACTING ANTIGEN NCA-95 (HUMAN);contains Alu repetitive element.
Score: 817 Identity: 433/443 (97%).
(22007..22267) AA32326 EST24948 Cerebellum II Homo sapiens cDNA 5' end. Score: 479 Identity: 252/262 (96%)."
repeat_region complement(21512..21626)
/rpt_family="Alu"
22506..22980
/note="DSS similarity to R88263 ym90h09.s1 Homo sapiens cDNA clone 166241 3'. Score: 837 Identity: 455/478 (95%).-Other overlapping matches:

(23013..22497) H05926 Y171f07.s1 Homo sapiens cDNA clone 43353 3'. Score: 872 Identity: 497/527 (94%).
(22971..22534) AA446916 zw85e01.s1 Soares total fetus NB2HF8 9w Homo sapiens cDNA clone 783720 3'. Score: 857 Identity: 436/437 (99%)
(22971..22561)W95490 ze02g07.r1 Soares fetal heart NBH119W Homo sapiens cDNA clone 357852 3'. Score: 774 Identity: 399/411 (97%).
(22566..22949) W95528 ze02g07.r1 Soares fetal heart NBH119W Homo sapiens cDNA clone 357852 5'.Score: 674 Identity: 366/382 (95%).
(22996..22634) T33668 EST58674 Homo sapiens cDNA 3' end similar to None..: Score: 617 Identity: 359/363 (98%)."
complement(23576..23770)
/note="DSS similarity to N94385 zb76u06.s1 Soares senescent fibroblasts NBHF Homo sapiens cDNA clone 309515 3'
similar to gb:219585 THROMBOSPONDIN 4 PRECURSOR (HUMAN); (194..1): 92% identity.-Other overlapping matches:
(23724..23770) predicted exon, program: grail2exons_human_1.3, frame: 1, quality: good, score: 74.000~(23770..23664) DDS similarity to w46954 mB82e10.r1 Soares mouse p3NMFI9.5 Mus musculus cDNA clone 335946 5' similar to
gb:219585 THROMBOSPONDIN 4 PRECURSOR (HUMAN): gb:M86620 Mouse thrombospondin 3 (mouse) (307..423): 63% identity."
complement(23683..32077)
/gene="COMP"
/note="COMP"
/gene="COMP"
complement(join(23724..23770,23683)..24002,25000..25172,25705..25901,26307..26355,26482..26660,26774..26955,27048..27100,27341..27458,28299..28458,28020..29127,29218..29322,29400..29558,29647..29721,29396..30105,30750..30922,31370..31421,31658..31743,31999..32077))
/gene="COMP"
/note="CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR (COMP)"
(COMP)"
/codon_start=1
/product="COMP_HUMAN"
/protein_id="AA86501.1"
/db_xref="GI:2623750"
/translation="MVPDFACVILLTLALAGSGGSGPLSGDLPGOMLREIQTNA LQVRELLROQVREITPLKTNVMECDCAQGSQSRVTLPSRYRLHARPGFPGVAC IQTESGARCGCPAGFTGNSHCTDVECNHNPCTFPVKCINTSPGRCAPGPGISG PTHQGVGLAFAPAKAQVCTDINECTGQHCNPNSVCINRSGQCGCPGPGDGA SGCORARQRCPCDPSPECHHADCVLERGSRCACVAGMAGNIGCRDIDLDFP DEKLRCERQCRKNCVTPVNSGQEDVRDGIIDAPCDPDAGDGVPRKNCPLVRNP DQRTDDEKMGKDCDNCRSOKNDQKOTDDIGDGDIDGRIINQDNCRVNP SDOKDSGDCIGDADCNCPKRSNPQADVHDHFDGACDSDQDDGSGHDSRNCPT VPNSAQEDSDHGGDADDNDGVPDSRDKLPNFGQEDADHDGVDGCDQDF DADRVPKIDVCPENAEVTLTDFRAFQTVLADPRGDAQIDPMWVLLNQGEKIVTQMS DPGILAVQYTFANGVDFGTGFHVNTDDVAGFFGQDSSFFVVMKMKMEQYTMAS NPFRAVVEPQIQKAVKSGTGPBQLRNALMHTGDTESQYRLMLKDDPRNVGMKDKSY

Query Match 0.88; Score 20; DB 85; Length 46275;
Best Local Similarity 100.0%; Freq. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

oy 1601 gggctctgggacacctggct 1620
|||||
Db 20016 gggctctgggacacctggct 20035

RESULT 28
AB007727 79590 bp DNA PLN 27-DEC-2000
LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, pl clone:MK9.
ACCESSION AB007727 BA000015
VERSION AB007727.1 GI:2696018
KEYWORDS Arabidopsis thaliana (strain:Columbia) DNA, clone_1lb:Mitsui P1 clone:MX9.

CDS	complement(24733..26592) /note="gene_id:MXC9.8 p1c1 T02643 strong similarity to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB10030.1" /db_xref="GI:9759379"
CDS	/translation="MSASAKMSKSDKRVNMDSQTPKASGSMGSSVYVYGVNPLIG FTQTESLSTGSSSLNNFRHIDSDSTGACDVSNNNGSGSEDEHKEVPT SLAKOETIGADNDKMRKLNKRNQKRAOELRRCQYLMKRLDIEVHOL IISMGIPHERATYALMNECKIESIMMLDDGAKYADKLPSTSGNLDLSQELG RILEETKYRATKODYERAVYTAEGDIEKEEALRROKODOSTRKRPEIDISTVN NSKVPSTLSTONVGOLOPNSGVPAGREALDRKNLNGPRSSYISGESNVSNT ERHMKLQMKLQONNAPLERNKRPYQOTPLSRSTETHYVTAIGDQFKRLQOQDME PVWMOQOQOQOSANTNVLPVSTNNSFTGAAGAGSGMHPNASEAQSNGYLP SRSLPNDLNSMLYQOLQYQOGVNNNGHMGAGSAPPAVAPASLILFSGFGL AGSSSGLDNNTGSLGHDYNNIDMSLDGLACVRSQOYVAASSPYASPYEAM MNGRTMTNGNGMGVGOEALVNGREMTSPFEGKDLFSLRQYVPSL" join(28348..28450,28599..28744,28826..28902,28993..29080, 29161..29241,29336..29392,29485..29566,29655..29708, 29854..29910,29996..30074,30190..30248,30406..30463, 30561..30658) /note="contains similarity to transmembrane transport protein gene_id:MXC9.9" /codon_start=1 /evidence=not_experimental /protein_id="BAB10031.1" /db_xref="GI:9759380" /translation="MLMCFWYTLISANRRLSHLPFASRGIDQDEDEKESRELLPH KNDENATSSSSSSVSGGLKDYQOETYSKTVACVGTANARIGIKREGYK ASEFPAGYILEGSLSDNLFVYLFYRYFVPLMYQNKVLTGAGATVRFILIG TALQKEAVNLIAAVLVSFLEFSEEDVDLSNFTVKTQRPITPVSYDGR FTFKHDIKATPPLLVAVIELSDIAVDSIPAVGVTRDPETIVLTSNFIALIGR SLYTLISGEGDELEYIQPSIAVYLGFIGVMKIIDFEGFHISTEASLGVALISTGV LSLTNKSQDS" complement(join(31799..31984,32321..32440)) /note="emb CA003929.1 gene_id:MXC9.10" /codon_start=1 /evidence=not_experimental /product="cystatin" /protein_id="BAB10032.1" /db_xref="GI:9759381" /translation="MADQAGTIVGVRIIDANANDLOVESLARFAVDENKKNLTL EYKRLGAKTQVAVGTMHMLTVEVADGETNKVYEAKLERAMEMLKOLESFNHLHY"
CDS	complement(join(3929.1..31984,32321..32440)) /note="emb CA003929.1 gene_id:MXC9.10" /codon_start=1 /evidence=not_experimental /product="cystatin" /protein_id="BAB10032.1" /db_xref="GI:9759381" /translation="MADQAGTIVGVRIIDANANDLOVESLARFAVDENKKNLTL EYKRLGAKTQVAVGTMHMLTVEVADGETNKVYEAKLERAMEMLKOLESFNHLHY"
Query Match	0.88; Score 20; DB 12; Length 79590; Best Local Similarity 100.0%; Pred. NO. 24; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1898 attggaagagctcgaagc 1917 Db 4335 atttggaagacttcagacc 4354
RESULT 29	AB013392 84129 bp DNA PLN 27-DEC-2000 LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone:M1K19. DEFINITION AB013392 BA000015 ACCESSION AB013392.1 GI:3128139 VERSION KEYWORDS SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1 clone:M1K19. ORGANISM Arabidopsis thaliana Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. REFERENCE 1 (sites) AUTHORS Kotani,H., Nakamura,Y., Sato,S., Asamizu,E., Kaneko,T., Miyajima,N. and Tabata,S.

TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. VI. Sequence features of the regions of 1,367,185 bp covered by 19 physically assigned P1 and T8C clones
JOURNAL	DNA Res. 5 (3), 203-216 (1998)
MEDLINE	98403884
REFERENCE	2 (bases 1 to 84129)
AUTHORS	Nakamura,Y.
TITLE	Direct Submission
JOURNAL	Submitted (06-MAY-1998) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934) Address for correspondence: kaos@kazusa.or.jp for the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/sgd/graph.cgi?c=M1K19 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Graft (Informatics Group, Oak Ridge National Laboratory, http://complib.ornl.gov/Graft-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremlin1.zool.iastate.edu/cgi-bin/ps.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MKN22 and the 3' clone is MP110. Location/Qualifiers 1..84129 /organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /chromosome="5" /clone="M1K19" /clone_lib="Mitsui P1" join(941..1855,1946..2080,2340..2609) /note="gene_id:M1K19.1 p1c1 T06019 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB09874.1" /db_xref="GI:10176760" /translation="MRKLFFPCTHVHVKORAKFTRIMESQOKKOLETPNELSMPPD LIRKIFSLPFRKDLATRFISRGKILNPDRAITDDSLMTCSFYVGSLSMRD AQILDSLHLKRYKTYASDINFLVQIDVNSMRRLIDFEGKTLFPCSLCTTTLK LVLDHLNIMSPGWFRLSVETLQSSVTGGSNVPSLIRGCSVHERLVVDONKRL VINVPTRLSLSDNKRGRHVPGLGSFWIINSLMFLIKIDTFPIESMAMIKANIE VAHDQOSVKEFIESLSTGRHLCSPTSENPYRGKTFESNVHLKCTGCSAGMONTLA CMINDANLSRLSTIKLRKSDVNPKNPKYKWEKPVVPELSTRLETKRDEGTGHEHD MGVTLANATFLQKATSTYKRNQCDSESELSQSMERYSEICEFVPD" complement(join(3056..3148,3239..3284,3384..3607, 3708..3890,3970..4197,4294..4428,4528..4612,4749..4825)) /note="gene_id:M1K19.2" /codon_start=1 /evidence=not_experimental /product="protein kinase Meki homolog" /protein_id="BAB09875.1" /db_xref="GI:10176761" /translation="MVKIKSNLKQLKLSVPAQESPISFTASGTFHDGDFLLNOKGL RLTSDKOSROSQSKELDPFTLADLTVAIVIGSGGVQVLYEVRHKKVGFPAKVIQ MNIDETRKQIVQELKYNQASQCPHYVNCYHSYHAGASIVLYEVRDGRSLATQ VKITLEPLVAVCKVILGLVYTHNEHNVYHTRDKPSNLLVNHKGVKISDFGVASL ASSMGQDRTFVGYVNVSPERISGTSYSSDIWSLQMSVLECAIGFPLYLESDQON PSPSEYELIAIVENPPTAPSDQFSPEFCFVSACIOKDPARASIDLILSHPIKKEF EDKIDIDGILVGTLEPPVYLR" join(5726..5780,5997..7280,7375..7408,7494..7641) /note="gene_id:M1K19.3"
FEATURES	
SOURCE	
CDS	
CDS	

/codon_start=1
/evidence=not_experimental
/product="beta-1,3-glucanase-like protein"
/protein_id="BAB09876.1"
/db_xref="GI:10176762"
/translation="MARDKRLFSISILLLLDCCYGVKAVCYGRSADLPTPSKV
LQIQNNIKYVRLYDYNQVLFKAFNTSIELMGPNSDLNAPSQSDVIMLKSV
LPYPTTKITVIVGAESIDDDPHINASSFVPMQVNLALRKVGLSRIRKYSSTLSL
GISRSFPPSAGAFSSVAYFLRPMLEPLAENKSPMIDLYPYVAYRSDPNNSLDYV
LPSSEVDPDNPNGLLXKNMFDQVAILALPMLRTIKMYETEGMPPTSGPKRE
MAASDNATETNSNITRHYVNOGTAKGGEAMNVIYFSLFNENKAGLDSENMWLF
YDQTSVYQDLDFGKSNFHSNNGSSGSSSWCIIASSKASRLKALWACAPG
MNDCTAIQSPQPCPDITLVSHASFVNSYFQGNRATDVACSFSGAGVYVNDPSDK
CIYTAGCKTKATNATATLSASSTRGNELLOMILKICIMISLFSLOTMSQAL"
complement(join(8087..8221,8288..8425,8793..8915))
/note="gene_id:MIK19.4"
/codon_start=1
/evidence=not_experimental
/product="profilin-like protein"
/protein_id="BAB09877.1"
/db_xref="GI:10176763"
/translation="MSMOTVYVDHLMCDVAGNRLTAAILIGDGSVMAQSNPNPYK
BEIQGKIDPFTPTGTLAPTGELFSGKKVWYQGEPAVAVIRGKAGAGVTIKTTLLV
FGIYDEPMTPGQCNVVENIDGTYLIESGL"
join(13604..13668,13748..14074,14151..14235)
/note="gene_id:MIK19.5"
/codon_start=1
/evidence=not_experimental
/product="refloement pol polyprotein-1-like"
/protein_id="BAB09878.1"
/db_xref="GI:10176764"
/translation="MTLYFIQKFEQKIMQYFQILSIYHHIODEDCMYAKKPTIM
LYENAVYIOLKQGYIKGDRTHILKPFYTHDLOKSGRVQIRLNNLIDFLFKA
LPAPRKILVYRGLKRCIKNIDYTHGHOHFYSGYGIWIRGVSNNKQCPPL"
join(14943..15139,15345..15429,15493..15582,15665..15758,
15841..15923,16024..16182)
/note="contains similarity to unknown protein
gblAD15447.1
gene_id:MIK19.6"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09879.1"
/db_xref="GI:10176765"
/translation="MTDEEDDTTOQRSSRNDGVSKNGKGFKGDKAKRALIAGGR
ILFYPILLNLVAFKLOSQFRCYADVRYNEFLPTKISEYQSGDFESMQKDPRLKL
GGVGYTLNPEYETLVPSLSYSAEMEHVIPRODLPASIVYDITLAVNFIHKMNL
GTTYVHCAGAGSTYVLYCILEHKSMTVAAPRHVSIRRVLLHFSQKRVSLDR
FIDMLQRDNRIYTKMKT"
complement(join(16726..16824,17130..17372,17821..18332,
18890..19082,19231..19344))
/note="gene_id:MIK19.7
p1r11713434
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09880.1"
/db_xref="GI:10176766"
/translation="MKNSKCNLIDSKLEEHNLGSGHCGGGMTOATKPNWGLP
AGVKEPDPDQLEHLEAKVYKGEKKSSSRILDEIPTIDGSDGCIYHNPOLP
GYTRDGLSKHFFHKPSRAYTGTGRKRIKLIQDHDSELIGSSSTRHMKGKTRPVIN
GQQRCKKILVLYTNFGKNRPKPTKMWVHQHLEINEERBELVSKIFQOTORQ
EASNTNMSDHGSKVDIGIGVDLSVATLQSLGSGVSVVNNHPTIRSFDECTA
CVSKRGNOHVGSTCEVHDGILITSSSMHMHIDHNOHNOHGDREFHMSYVMT
PITTSOHEISIFVTSITPFGSGVYMGIRPNVGLYHSVYRRRKRLTNLCY"
join(22516..22617,22808..22854,22935..22974,23115..23324,
23408..23566,23656..23709,23800..23998,24082..24344,
24431..24631,24742..24933)
/note="gene_id:MIK19.8"
/codon_start=1
/evidence=not_experimental
/product="pyrophosphate-dependent
phosphotruco-1-kinase-like protein"
/protein_id="BAB09881.1"

/db_xref="GI:10176767"
/translation="MSSPRSNKPKIYVNGGXYILODPHLIDYLPDLPTYPNPLODP
AYSVYKQYVHADDSVPEKRVVYHKGSPRGVHRRRAGPRQKVFESSEVACIVTGGGL
CEGLANVIREVYSSLSYMGVARKILGIDVSLGSGYREFYAKNTIPILNSKVVNDHRRGG
TIIIGTSRGHDHDKIVDSIQDKGINOVYIIGDGOYKRGASVFEELRRRLKLVAVGI
PTIIDNDIPVINDKSEFEDPAVEAQAIAVAHVEASNSNGIGEPKLMERYSGYIAMY
ATLASHDYVCCILIPESFYLEGGGLFEETIERRLKPHGMIVLVLEGAQDMLCSME
SPMDASGKLKLDKVLMSQSIKHDFKNNKMVNLKYIDPTYMIKAVPSNADNYC
TLAOSAVHMGAGYGTSGLVNGROYIIPYRITETRONNVITDRMARRLLSSTNQ
PSTFGKDISEAKKEKLEPPLDGDGVDIIPYKTEYTK"
complement(join(25739..25765,25861..25942,26048..26126,
26210..26279,26372..26494,26652..26704,26807..26866,
26966..27110,27390..27474,27561..27690,27813..27843))
/note="gene_id:MIK19.9
p1r11706010
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09882.1"
/db_xref="GI:10176768"
/translation="MNISVENPVFVHEDSTQKTGELRLDSIDIPMSKISSDDDEVFLAP
ENANREORFQDTPDTPNSERKSEVHEVYATQHNQTLDPQKMRSEGKIDKVMNWE
CELSKEVYDESDDIDEPQIQLDQSAEAIKRDVNEWMILRALIIDLKGVLTLPQ
FGGLPQMAVVGDTFVPGCAFDESNNHKKFEMNPENRKYTKAGIYSBGGLENVL
MSGHDHYMYLVAKENGSTLPSGLFTIIRHSEFYPLHKKAGAYTHLMNEDEKRLWLH
VFNKYDLYSKSVYHVEKRYMGLIKKYPENLRW"
join(29030..29383,29492..29617,29692..30006,30088..30210,
30295..30693)
/note="gene_id:MIK19.10
sp1P54969"
/codon_start=1
/evidence=not_experimental
/product="IAA-amino acid hydrolase homolog 1 precursor"
/protein_id="BAB09883.1"
/db_xref="GI:10176769"
/translation="MALNMFILTFQILLILRVSSSPWIVAGDVSRIPINLELAKSP
EYFDSNVRIRRKITHENPELGYEFETSKIRBELDLYGVYFPVAITGIIDYISGGE
PPVVALRADMDALPIQEAWEHKSINPKAGCDHGVAMILGKILQOHRQHL
GTVVLIPOAEGLSGAKMREGALKNEVLEFGLHSRTPFGKASLAGSFMAGAG
AFEAIVITGKGHAIPQHTIDPVVAASIVLQLHYSSETPSPDSKVYTVKVNCGN

Query Match 0.8%; Score 20; DB 12; Length 84129;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1703 cttgggaagccgcttcac 1722
DB 3235 ctttgggaagccgcttcac 32354

RESULT 30
ATF14F18
LOCUS ATF14F18 103638 bp DNA PLN 13-APR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 5, BAC clone f14f18 (ESSA
project).
ACCESSION AL163812
VERSION AL163812.1 GI:7573350
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 103638)
Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A.,
Duesterhoelt,A., Bancroft,I., Mewes,H.W., Rudd,S., Lemcke,K. and
Mayer,K.F.X.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 103638)
EU Arabidopsis sequencing, project.
TITLES Direct Submission
JOURNAL Submitted (13-APR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitzz 18a, D-82152 Martinsried, FRG, E-mail:

1emcke@mps.biochem.mpg.de,mayer@mps.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.
Location/Qualifiers
1. 103638
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
1. 8764
/note="overlap with BAC T22P22. Please refer to this BAC
for annotation and analysis"
complement(8151..8375)
/gene="F14F18_10"
/number=1
complement(8151..8375,8495..8551,8650..8678,
8890..9346)
/gene="F14F18_10"
8151..9346
/gene="F14F18_10"
complement(101n(8151..8375,8495..8551,8650..8678,
8890..9346))
/gene="F14F18_10"
/note="similarity to predicted proteins in *Synechococcus*,
Cyanophora and other organisms"
/codon_start=1
/product="putative protein"
/protein_id="CA887657.1"
/db_xref="GI:7573351"
/translation="MAATVGHPSPTTSLKPPSRHRLAHNLSPKTPLOSSLIRKG
AENSEQPPRPLEPDTCPVPEQOPINEXSLSTSPFWSGDLLEYSTRFLTGAS
FAFVGLPVSMFGSIGVEYKRIILAASSGIFVTVLAVRMVGLGAYDEEGWTD
GOVWKTPEVLARDRLGSESVKPVILARKNTLVILGSLILVINLGDSPILAIYSKRY
RPPDRSSSLPIGAYNDERTARTEPPAFCEPSSDL"
complement(8376..8494)
/number=1
complement(8495..8551)
/gene="F14F18_10"
/number=2
complement(8552..8649)
/number=2
complement(8650..8678)
/gene="F14F18_10"
/number=3
complement(8679..8889)
/number=3
complement(8890..9346)
/gene="F14F18_10"
/number=4
complement(101n(11457..11637,11711..11760,11866..12033,
12133..12211,12407..12507,12606..12674,12823..12863,
13089..13356,13438..14056,14134..14237,14416..14856,
14953..15048,15122..15565))
/gene="F14F18_20"
complement(11457..11637)
/gene="F14F18_20"
/number=1
11457..15565
/gene="F14F18_20"
complement(101n(11457..11637,11711..11760,11866..12033,
12133..12211,12407..12507,12606..12674,12823..12863,
13089..13356,13438..14056,14134..14237,14416..14856,
14953..15048,15122..15565))
/gene="F14F18_20"
/note="The mRNA covering this gene is only partial. This
has allowed the definition of the carboxy terminus of the
protein. The amino terminus remains predicted but
prediction is supported by different gene models."
/codon_start=1

/product="MAP3k delta-1 protein kinase"
/protein_id="CA887658.1"
/db_xref="GI:7573352"
/translation="MSRKHLLKRLHIGSSGVGGGFADHRLDSTRMIDPSPLIS
TSPSPASTSSSSGFGNASTMPRLDTPFVGRLTAADGVGFNMLEEYQVQLMA
ISVSDPPRENADTAQIDAKRISLSASAVTDADSAVDLSLRMSGCSISGLTEL
GSMILPAVFSSILRFLNPGCHGKVIYNDKVRDGFYVYGITNSLSOGMPLVDLO
AISTIDNVDEVLIVNRLIDPELOELRRVRFALASCPDPAPOVSDLLTKANTIV
EQMGCPENADEALRRMLRSYELRNSLNTIILPLRVNGLARHRLLEKVLADRIN
LPCLVKGSTYTGTGDAVNLIKLDKSPNSFTCSBYIIDLGAFCALIPSEVPSSF
LPVSCDTRVFPENLDSLQHSVLEKEIETPAFVSKEADSSGMVANFTGNOEN
SDRCVAKKHQTERFEHDFGKIMHSQOISGMNPPESGKPTCAQKVKVNSKVISKA
KNPEFAOKHVAVLLESGASPPDLFMDINPHNLRGKMLDELROESSNVSGLPCPV
EKVADELRESENRPTAESYOOSEVIDSMKRNPDLDNTGKASSSENNEVGTAGESAV
CDSHDQINPLLEBAKWEIMWEDLQIGERIGSGISGEYIRAEKNGTEVAVKFFLDD
FSGDALTFQKSEIEIMRLRHPNVVLEMGAVTPPNPSILTEFLPRGSLRLRHPNH
QLDEKRRMRMALDVAKGMNLTLSHPVVRHDLKSPMLVDKMMVVKVCDPGLSRMKH
HTYLSKSTAGTEPMMAPEVLRRNEPAEKCDVYSFGVILWELATSRVPMKGLNMQVY
GAVGFORRLREIPDDIDLTFVAQIIRBCWQT"
complement(11638..11710)
/number=1
complement(11711..11760)
/gene="F14F18_20"
/number=2
complement(11761..11865)
/number=2
complement(11866..12033)
/gene="F14F18_20"
/number=3
complement(12034..12132)
/number=3
complement(12133..12211)
/gene="F14F18_20"
/number=4
complement(12212..12406)
/number=4
complement(12407..12507)
/gene="F14F18_20"
/number=5
complement(12508..12605)
/number=5
complement(12606..12674)
/gene="F14F18_20"
/number=6
complement(12675..12822)
/number=6
complement(12823..12863)
/gene="F14F18_20"
/number=7
complement(12864..13088)
/number=7
complement(13089..13356)
/gene="F14F18_20"
/number=8
complement(13357..13437)
/number=8
complement(13438..14056)
/gene="F14F18_20"
/number=9
complement(14057..14133)
/number=9
complement(14134..14237)
/gene="F14F18_20"
/number=10
complement(14238..14415)
/number=10
complement(14416..14856)
/gene="F14F18_20"
/number=11
complement(14857..14952)
/number=11
complement(14953..15048)
/gene="F14F18_20"


```

repeat_region 16066..16225
/note="LIME repeat: matches 420..587 of consensus"
repeat_region 16706..17508
/note="LIME3 repeat: matches 5198..6013 of consensus"
repeat_region 17677..17818
/note="L1 repeat: matches 4558..4701 of consensus"
repeat_region 17834..18129
/note="AluSc repeat: matches 1..295 of consensus"
repeat_region 18378..19341
/note="L1PB8 repeat: matches 5187..6163 of consensus"
repeat_region 19343..19406
/note="32 copies 2 mer aa 76% conserved"
repeat_region 19407..19665
/note="AluY repeat: matches 53..311 of consensus"
repeat_region 20762..20885
/note="MIR repeat: matches 113..232 of consensus"
repeat_region 22524..23654
/note="Tiger3b repeat: matches 1..1203 of consensus"
repeat_region 25639..25680
/note="21 copies 2 mer aa 76% conserved"
repeat_region 26701..26842
/note="71 copies 2 mer aa 60% conserved"
repeat_region 27744..28097
/note="MLT1B repeat: matches 1..554 of consensus"
repeat_region 28204..28331
/note="64 copies 2 mer aa 60% conserved"
repeat_region 28716..29064
/note="LIM4 repeat: matches 2557..2927 of consensus"
repeat_region 29065..29367
/note="AluJo repeat: matches 1..296 of consensus"
repeat_region 29368..29387
/note="LIM4 repeat: matches 2927..2945 of consensus"
repeat_region 29545..29596
/note="LIM4 repeat: matches 3172..3223 of consensus"
repeat_region 29614..29756
/note="LIME1 repeat: matches 5810..5950 of consensus"
repeat_region 31279..31589
/note="AluSc repeat: matches 1..313 of consensus"
repeat_region 31911..32226
/note="AluJb repeat: matches 1..307 of consensus"
repeat_region 32289..32452
/note="L1PA13 repeat: matches 5992..6156 of consensus"
repeat_region 36609..37560
/note="Cpg island"
/evidence=not_experimental
repeat_region 38216..38375
/note="MER91A repeat: matches 3..173 of consensus"
repeat_region 38751..38782
/note="16 copies 2 mer ca 100% conserved"
repeat_region 40777..40946
/note="MER5B repeat: matches 1..172 of consensus"
repeat_region 41788..42159
/note="LIMC5 repeat: matches 7311..7699 of consensus"
repeat_region 43207..43513
/note="AluY repeat: matches 1..303 of consensus"
repeat_region 43641..43764
/note="FLM_C repeat: matches 1..125 of consensus"
repeat_region 44458..44447
/note="L1PA15 repeat: matches 1396..5414 of consensus"
repeat_region 48448..48590
/note="AluJb repeat: matches 1..150 of consensus"
repeat_region 48591..48890
/note="AluSc repeat: matches 1..306 of consensus"
repeat_region 48891..49074
/note="AluJb repeat: matches 150..310 of consensus"
repeat_region 49075..49819
/note="L1PA15 repeat: matches 5414..6151 of consensus"
repeat_region 50221..50406
/note="L2 repeat: matches 2529..2696 of consensus"
repeat_region 50407..50716
/note="AluY repeat: matches 1..304 of consensus"
repeat_region 50717..50823
/note="L2 repeat: matches 2433..2529 of consensus"

```

```

repeat_region 50824..51118
/note="AluYa5 repeat: matches 1..299 of consensus"
repeat_region 51119..51171
/note="L2 repeat: matches 2270..2433 of consensus"
repeat_region 52154..52578
/note="MSTRB repeat: matches 1..426 of consensus"
repeat_region 52844..53363
/note="MER44B repeat: matches 10..530 of consensus"
repeat_region 54572..54751
/note="MER5A repeat: matches 4..186 of consensus"
repeat_region 55699..56343
/note="L1PB2 repeat: matches 5493..6152 of consensus"
repeat_region 58368..58536
/note="THER1B repeat: matches 1..165 of consensus"
repeat_region 58537..58827
/note="AluSc repeat: matches 1..291 of consensus"
repeat_region 58828..59013
/note="THER1B repeat: matches 165..364 of consensus"
repeat_region 59014..60590
/note="THER1B-INTERNAL repeat: matches 1..1580 of consensus"
repeat_region 60591..60668
/note="THER1B repeat: matches 1..93 of consensus"
repeat_region 60670..60792
/note="THER1-INTERNAL repeat: matches 1060..1182 of consensus"

```

```

Query Match 0.88; Score 20; DB 90; Length 114521;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2272 ccacaaatgcccaagctgat 2291
Db 47417 CCAACAATGCCCAAGCTGAT 47436

```

```

RESULT 32
HS173D1/C
LOCUS HS173D1
DEFINITION Human DNA sequence from clone 173D1 on chromosome 1p36.21-36.33.
ACCESSION AL031984
VERSION AL031984.13 GI:4678432
KEYWORDS HNG; Cpg Island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE Moore,M.
AUTHORS Moore,M.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Apr 24, 1999 this sequence version replaced g1:4584747.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em; EMBL; Sw; SWISSPROT; Tr; TREMBL.
This sequence is the entire insert of clone 173D1. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1

```

Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Ch117301> is from the library RPC1 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/VECTOR>: PCIPAC2.

FEATURES

source

```
1. .117338
/organism="Homo sapiens"
/db xref="taxon:9606"
/chromosome="1"
/map="p36.21-36.33"
/clone="RP1-173D1"
/clone_lib="RPC1-1"
91. .142
/note="13 copies 4 mer aggg 75% conserved"
repeat_region
585. .992
/note="17 copies 24 mer 97% conserved"
repeat_region
1728. .1903
/note="MER45 repeat: matches 1. .178 of consensus"
repeat_region
1909. .2078
/note="MIR repeat: matches 28. .187 of consensus"
repeat_region
4058. .4357
/note="AluY repeat: matches 1. .298 of consensus"
repeat_region
4464. .4571
/note="2 copies 54 mer 96% conserved"
repeat_region
4726. .4844
/note="MIR repeat: matches 21. .151 of consensus"
repeat_region
5494. .5786
/note="AluSg repeat: matches 1. .292 of consensus"
repeat_region
6503. .6807
/note="AluJo repeat: matches 1. .311 of consensus"
repeat_region
6809. .6857
/note="MIR repeat: matches 101. .150 of consensus"
repeat_region
6858. .7181
/note="AluX repeat: matches 1. .311 of consensus"
repeat_region
7182. .7299
/note="MIR repeat: matches 150. .262 of consensus"
repeat_region
8932. .8993
/note="MER53 repeat: matches 126. .189 of consensus"
repeat_region
8994. .9305
/note="AluSg repeat: matches 1. .311 of consensus"
repeat_region
9306. .9440
/note="MER53 repeat: matches 1. .126 of consensus"
repeat_region
10094. .10654
/note="12 repeat: matches 2152. .2746 of consensus"
repeat_region
13515. .13745
/note="12 repeat: matches 1886. .2084 of consensus"
repeat_region
13746. .14051
/note="AluX repeat: matches 1. .306 of consensus"
repeat_region
14052. .14530
/note="12 repeat: matches 2084. .2680 of consensus"
repeat_region
14561. .14871
/note="AluX repeat: matches 1. .311 of consensus"
repeat_region
16630. .16703
/note="MIR repeat: matches 101. .176 of consensus"
misc_feature
complement(17036. .17418)
/note="match: STS G11038"
19534. .19676
/note="MIR repeat: matches 26. .168 of consensus"
repeat_region
19811. .20021
/note="12 repeat: matches 2521. .2705 of consensus"
repeat_region
20103. .20402
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region
20551. .20620
/note="12 repeat: matches 2632. .2701 of consensus"
repeat_region
20681. .20878
/note="MIR repeat: matches 28. .238 of consensus"
repeat_region
22445. .22468
/note="12 copies 2 mer ca 100% conserved"
repeat_region
22496. .22714
/note="MIR repeat: matches 2. .223 of consensus"
repeat_region
22969. .23143
/note="MIR repeat: matches 71. .232 of consensus"
```

```
repeat_region
23333. .23460
/note="2 copies 64 mer 98% conserved"
repeat_region
24447. .24757
/note="AluSg repeat: matches 1. .309 of consensus"
repeat_region
25099. .25400
/note="AluX repeat: matches 1. .303 of consensus"
repeat_region
25523. .25637
/note="Char1e5 repeat: matches 5. .121 of consensus"
repeat_region
25767. .26056
/note="AluX repeat: matches 1. .291 of consensus"
repeat_region
26059. .26270
/note="MIR repeat: matches 35. .256 of consensus"
repeat_region
26313. .26624
/note="AluY repeat: matches 1. .312 of consensus"
repeat_region
27898. .27939
/note="7 copies 6 mer tcttc 95% conserved"
repeat_region
29266. .29305
/note="10 copies 4 mer aagg 80% conserved"
misc_feature
31618. .31978
/note="match: EST A1350672"
repeat_region
33293. .33452
/note="MIR repeat: matches 48. .203 of consensus"
repeat_region
33832. .34011
/note="90 copies 2 mer ac 71% conserved"
repeat_region
33832. .33999
/note="7 copies 24 mer 73% conserved"
repeat_region
33832. .34023
/note="3 copies 64 mer 72% conserved"
repeat_region
33832. .33993
/note="3 copies 54 mer 74% conserved"
repeat_region
34018. .34503
/note="9 copies 54 mer 67% conserved"
repeat_region
34018. .34497
/note="20 copies 24 mer 66% conserved"
repeat_region
34018. .34505
/note="122 copies 4 mer caca 65% conserved"
repeat_region
34018. .34491
/note="237 copies 2 mer ca 66% conserved"
repeat_region
34025. .34504
/note="48 copies 10 mer acacacacac 65% conserved"
repeat_region
34028. .34507
/note="80 copies 6 mer cacaca 65% conserved"
repeat_region
34065. .34512
/note="7 copies 64 mer 70% conserved"
repeat_region
34501. .34914
/note="3 copies 138 mer 70% conserved"
repeat_region
34508. .34837
/note="33 copies 10 mer acacacacac 66% conserved"
repeat_region
34510. .34839
/note="55 copies 6 mer acacac 66% conserved"
repeat_region
34511. .34846
/note="14 copies 24 mer 66% conserved"
repeat_region
34519. .34842
/note="6 copies 54 mer 67% conserved"
repeat_region
34519. .34838
/note="160 copies 2 mer ca 66% conserved"
repeat_region
35512. .35541
/note="15 copies 2 mer tt 90% conserved"
misc_feature
complement(36540. .36839)
/note="match: EST AA295567"
36962. .37180
/note="MIR repeat: matches 9. .251 of consensus"
repeat_region
37142. .37185
/note="12 repeat: matches 2655. .2698 of consensus"
repeat_region
37989. .38163
/note="MIR repeat: matches 47. .236 of consensus"
repeat_region
39487. .39536
/note="5 copies 10 mer ggtgtgtgt 90% conserved"
repeat_region
39487. .39528
/note="21 copies 2 mer gt 98% conserved"
repeat_region
39487. .39534
/note="12 copies 4 mer gttgt 92% conserved"
repeat_region
39831. .40133
```



```
repeat_region /note="AluX repeat: matches 1. .304 of consensus"
40195. .40373
/note="MIR repeat: matches 32. .222 of consensus"
repeat_region 40374. .40710
/note="MER3 repeat: matches 9. .149 of consensus"
repeat_region 40684. .40748
/note="Charlie5 repeat: matches 1. .62 of consensus"
repeat_region 40795. .41105
/note="AluX repeat: matches 1. .309 of consensus"
misc.feature complement(41047. .41397)
/note="match: STS L31008"
repeat_region 41244. .41387
/note="72 copies 2 mer to 63% conserved"
repeat_region 41263. .41538
/note="46 copies 6 mer tttctt 57% conserved"
repeat_region 41264. .41527
/note="11 copies 24 mer 61% conserved"
repeat_region 41271. .41465
/note="5 copies 39 mer 73% conserved"

Query Match 0.88; Score 20; DB 92; Length 117338;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1149 caagatcaaccagctca 1168
|||||
Db 68061 CAAAGTTCACCAAGCTCA 68042

RESULT 33
AC022835
LOCUS Homo sapiens chromosome 15 clone RP11-15513 map 15, LOW-PASS
SEQUENCE SAMPLING.
AC022835
AC022835.2 GI:7767812
HTG: HTGS_PHASE0.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 127953)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 15, clone RP11-15513.
REFERENCE 2 (bases 1 to 127953)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguski,K., Bouckhalter,B., Brown,A., Burkett,G., Castle,A.,
Chen,Y., Collins,S., Collins,S., Collymore,A., Cooke,P.,
DeRubeis,K., Dewar,K., Domino,M., Doyle,M., Fennell,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heath,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lander,E., Lechekzy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Melgrim,J., Menus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 10, 2000 this sequence version replaced gi:6922258.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
```

```
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6143
Center clone name: 155_L_3
-----
* NOTE: This record contains 141 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 754: contig of 754 bp in length
* 755 854: gap of 100 bp
* 855 1641: contig of 787 bp in length
* 1642 1741: gap of 100 bp
* 1742 2528: contig of 787 bp in length
* 2529 2628: gap of 100 bp
* 2629 3393: contig of 765 bp in length
* 3394 3493: gap of 100 bp
* 3494 4259: contig of 766 bp in length
* 4260 4359: gap of 100 bp
* 4360 5136: contig of 777 bp in length
* 5137 5236: gap of 100 bp
* 5237 6004: contig of 768 bp in length
* 6005 6104: gap of 100 bp
* 6105 6897: contig of 793 bp in length
* 6898 6997: gap of 100 bp
* 6998 7757: contig of 760 bp in length
* 7758 7857: gap of 100 bp
* 7858 8626: contig of 769 bp in length
* 8627 8726: gap of 100 bp
* 8727 9505: contig of 779 bp in length
* 9506 9605: gap of 100 bp
* 9606 10390: contig of 785 bp in length
* 10391 10490: gap of 100 bp
* 10491 11261: contig of 771 bp in length
* 11262 11361: gap of 100 bp
* 11362 12122: contig of 761 bp in length
* 12123 12222: gap of 100 bp
* 12223 13003: contig of 781 bp in length
* 13004 13103: gap of 100 bp
* 13104 13869: contig of 766 bp in length
* 13870 13969: gap of 100 bp
* 13970 14741: contig of 772 bp in length
* 14742 14841: gap of 100 bp
* 14842 15621: contig of 780 bp in length
* 15622 15721: gap of 100 bp
* 15722 16478: contig of 757 bp in length
* 16479 16578: gap of 100 bp
* 16579 17361: contig of 783 bp in length
* 17362 17461: gap of 100 bp
* 17462 18244: contig of 783 bp in length
* 18245 18344: gap of 100 bp
* 18345 19111: contig of 767 bp in length
* 19112 19211: gap of 100 bp
* 19212 19977: contig of 766 bp in length
* 19978 20077: gap of 100 bp
* 20078 20846: contig of 769 bp in length
* 20847 20946: gap of 100 bp
* 20947 21718: contig of 772 bp in length
* 21719 21818: gap of 100 bp
* 21819 22577: contig of 759 bp in length
* 22578 22677: gap of 100 bp
* 22678 23409: contig of 732 bp in length
* 23410 23509: gap of 100 bp
* 23510 24277: contig of 768 bp in length
* 24278 24377: gap of 100 bp
```

```

* 24378 25142: contig of 765 bp in length
* 25143 25242: gap of 100 bp
* 25243 26013: contig of 771 bp in length
* 26014 26113: gap of 100 bp
* 26114 26886: contig of 773 bp in length
* 26887 26986: gap of 100 bp
* 26987 27736: contig of 750 bp in length
* 27737 27836: gap of 100 bp
* 27837 28594: contig of 758 bp in length
* 28595 28694: gap of 100 bp
* 28695 29481: contig of 787 bp in length
* 29482 29581: gap of 100 bp
* 29582 30356: contig of 775 bp in length
* 30357 30456: gap of 100 bp
* 30457 31224: contig of 768 bp in length
* 31225 31324: gap of 100 bp
* 31325 32094: contig of 770 bp in length
* 32095 32194: gap of 100 bp
* 32195 32984: contig of 790 bp in length
* 32985 33084: gap of 100 bp
* 33085 33851: contig of 767 bp in length
* 33852 33951: gap of 100 bp
* 33952 34708: contig of 757 bp in length
* 34709 34808: gap of 100 bp
* 34809 35590: contig of 782 bp in length
* 35591 35690: gap of 100 bp
* 35691 36455: contig of 765 bp in length
* 36456 36555: gap of 100 bp
* 36556 37323: contig of 768 bp in length
* 37324 37423: gap of 100 bp
* 37424 38185: contig of 762 bp in length
* 38186 38285: gap of 100 bp
* 38286 39061: contig of 776 bp in length
* 39062 39161: gap of 100 bp
* 39162 39930: contig of 769 bp in length
* 39931 40030: gap of 100 bp
* 40031 40787: contig of 757 bp in length
* 40788 40887: gap of 100 bp
* 40888 41662: contig of 775 bp in length
* 41663 41762: gap of 100 bp
* 41763 42512: contig of 750 bp in length
* 42513 42612: gap of 100 bp
* 42613 43386: contig of 774 bp in length
* 43387 43486: gap of 100 bp
* 43487 44254: contig of 768 bp in length
* 44255 44354: gap of 100 bp
* 44355 45116: contig of 762 bp in length
* 45117 45216: gap of 100 bp
* 45217 45977: contig of 761 bp in length
* 45978 46077: gap of 100 bp
* 46078 46847: contig of 770 bp in length
* 46848 46947: gap of 100 bp
* 46948 47712: contig of 765 bp in length
* 47713 47812: gap of 100 bp
* 47813 48582: contig of 770 bp in length
* 48583 48682: gap of 100 bp
* 48683 49462: contig of 780 bp in length
* 49463 49562: gap of 100 bp
* 49563 50335: contig of 773 bp in length
* 50336 50435: gap of 100 bp
* 50436 51202: contig of 767 bp in length
* 51203 51302: gap of 100 bp
* 51303 52119: contig of 817 bp in length
* 52120 52219: gap of 100 bp
* 52220 53062: contig of 843 bp in length
* 53063 53162: gap of 100 bp
* 53163 53995: contig of 833 bp in length
* 53996 54095: gap of 100 bp
* 54096 54925: contig of 830 bp in length
* 54926 55025: gap of 100 bp
* 55026 55884: contig of 859 bp in length
* 55885 55984: gap of 100 bp
* 55985 56818: contig of 834 bp in length

```

```

Query Match
Best Local Similarity 100.0%; Score 20; DB 67; Length 127953;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2447 agagccacagccagaag 2466
Db 69843 AGGAGCCACAGCCAGAAG 69862
|||||

```

```

RESULT 34
AC027795/C
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-235B15 map 18, WORKING DRAFT
SEQUENCE
AC027795
VERSION
AC027795.2 GI:7684508
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

REFERENCE
1 (bases 1 to 148508)
Birken, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 18, clone RP11-235B15
Unpublished
2 (bases 1 to 148508)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,
Boguslavsky, L., Bouknight, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, U., Lacombe, K., Lamazares, R., Landers, T., Lehotzky, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPheters, R.,
Meldrum, J., Menus, L., Mihova, T., Miyanaga, C., Miyanaga, Y., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testafave, S., Theodore, J., Tirrell, A., Travers, M., Wu, X., Wyman, D., Ye, W.J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

```

```

TITLE
JOURNAL
COMMENT
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information

```

```

Center project name: L9034
Center clone name: 235_B.15
----- Summary Statistics -----
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 138322 bases at least Q40
Consensus quality: 143906 bases at least Q30
Consensus quality: 145997 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 146908; sum-of-ctrls
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1262: contig of 1262 bp in length
* 1263 1362: gap of 100 bp
* 1363 2466: contig of 1104 bp in length
* 2467 2566: gap of 100 bp
* 2567 3867: contig of 1301 bp in length
* 3868 3967: gap of 100 bp
* 3968 4372: contig of 405 bp in length
* 4373 4472: gap of 100 bp
* 4473 7936: contig of 3464 bp in length
* 7937 8036: gap of 100 bp
* 8037 11973: contig of 3937 bp in length
* 11974 12073: gap of 100 bp
* 12074 16804: contig of 4731 bp in length
* 16805 16904: gap of 100 bp
* 16905 22543: contig of 5639 bp in length
* 22544 22643: gap of 100 bp
* 22644 27921: contig of 5278 bp in length
* 27922 28021: gap of 100 bp
* 28022 33883: contig of 5862 bp in length
* 33884 33983: gap of 100 bp
* 33984 39792: contig of 5809 bp in length
* 39793 39892: gap of 100 bp
* 39893 49448: contig of 9556 bp in length
* 49449 49548: gap of 100 bp
* 49549 59577: contig of 10029 bp in length
* 59578 59677: gap of 100 bp
* 59678 69844: contig of 10167 bp in length
* 69845 69944: gap of 100 bp
* 69945 95609: contig of 25665 bp in length
* 95610 95709: gap of 100 bp
* 95710 120155: contig of 24446 bp in length
* 120156 120255: gap of 100 bp
* 120256 148508: contig of 28253 bp in length.
FEATURES
Source
1. 148508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-235B15"
/clone_lib="RPC1-11 Human Male BAC"
1. 1262
/feature="assembly_fragment"
1363..2466
/feature="assembly_fragment"
2567..3867
/feature="assembly_fragment"
3968..4372
/feature="assembly_fragment"
clone_end:SP6
vector_side:right"

```

```

misc_feature 4473..7936
/feature="assembly_fragment"
misc_feature 8037..11973
/feature="assembly_fragment"
misc_feature 12074..16804
/feature="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature 16905..22543
/feature="assembly_fragment"
misc_feature 22644..27921
/feature="assembly_fragment"
misc_feature 28022..33883
/feature="assembly_fragment"
misc_feature 33984..39792
/feature="assembly_fragment"
misc_feature 39893..49448
/feature="assembly_fragment"
misc_feature 49549..59577
/feature="assembly_fragment"
misc_feature 59678..69844
/feature="assembly_fragment"
misc_feature 69945..95609
/feature="assembly_fragment"
misc_feature 95710..120155
/feature="assembly_fragment"
misc_feature 120256..148508
/feature="assembly_fragment"
BASE COUNT 44364 a 29387 c 30187 g 42968 t 1602 others
ORIGIN
Query Match 0.88; Score 20; DB 71; Length 148508;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1443 agaatcatctctcttgaa 1462
|||||
Db 9577 AGAATCATCTCTCTTGAA 9558
-----
RESULT 35
AC025812 149006 bp DNA PRI 03-OCT-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-182G24, complete sequence.
ACCESSION AC025812
VERSION AC025812.11 GI:10518383
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 149006)
REFERENCE 1
Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Meo,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
Ramirez,D., Wilhelm,J., Yu,S. and Davis,R.W.
Unpublished
2 (bases 1 to 149006)
JOURNAL Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D., Federspiel,N.,
Glukhov,S., Hansen,N., Hyman,R., Meo,J., Marathe,R.,
Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D.,
Wilhelm,J., Yu,S. and Davis,R.W.
Direct Submission
TITLE Submitted (15-MAR-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
(bases 1 to 149006)
REFERENCE 3
Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Meo,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
Ramirez,D., Wilhelm,J., Yu,S. and Davis,R.W.
Direct Submission
TITLE Direct Submission

```


misc_feature vector_side:right
40398..59678
/note="assembly_fragment"
misc_feature 59779..102333
/note="assembly_fragment"
clone_end:SP6
misc_feature vector_side:right
102434..149180
/note="assembly_fragment"
BASE COUNT 42836 a 32086 c 31253 g 42201 t 804 others
ORIGIN

Query Match 0.8%; Score 20; DB 63; Length 149180;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1588 gaccagctgacagagctcct 1607
|||||
Db 135751 GACCAGCTGACAGGCTCCT 135770

RESULT 37
AC069420 150683 bp DNA HTG 11-SEP-2000
LOCUS Homo sapiens chromosome 3 clone RP11-683J13, WORKING DRAFT
DEFINITION
SEQUENCE, 23 unordered pieces.
AC069420
AC069420 6 GI:10047548
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 150683)
Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,
Albrooks S.L., Amaralunga H.C., Are J.R., Banks T., Barbara J.,
Benton J., Bimaga K., Blankenburg K., Bonnin D., Bouck J.,
Bowie S., Brieve M., Brown E., Brown M., Bryant N.P., Bunay C.,
Butch P., Burkett C., Burrell K.L., Byrd N.C., Cartron T.F.,
Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R.,
Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C.,
Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,
Davy Carroll L., Dederich D.A., Delaney K.R., Delgado O.,
Denn A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H.,
Dugan-Rocha S., Durbin K.J., Earhart C., Edgar D., Edwards C.C.,
Elhaj C., Escotto M., Falls T., Ferraguto D., Flaeg N., Ford J.,
Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T.,
Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S.,
Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hawes A.,
Hernandez J., Hernandez O., Hodgson A., Hognes M., Holloway C.,
Hollins B., Homsl F., Howard S., Huber J., Huylk S., Hume J.,
Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S.,
Joudah S., Karlsson E., Kelly S., Khan U., King L., Koryah J.,
Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C.,
Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu M.,
Luissegue H., Lozano R.J., Lu X., Lucier A., Lucier R., Luna R.,
Ma J., Maheshwari M., Mapua P., Martin R., Martindale A.,
Martinez E., Massey E., Mawhney E., McLeod M.P., Meador M.,
Mei G., Metzker M., Miner G., Miner Z., Mitchell T., Mohabbat K.,
Morgan M., Morris S., Moser M., Neal D., Newton J., Newton N.,
Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nokenkwo S.,
Ogutu M., Okwuonu G., Otagunye N., Oyedero R., Pace A., Payton B.,
Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L.,
Ruiz S., Savary G., Scherer S., Scott G., Shen H., Shoshitani N.,
Slisom I., Sodergren E., Sonakke T., Sparks A., Stanley H.,
Stone H., Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K.,
Tang H., Tansey J., Taylor C., Taylor T., Telford B., Thomas N.,
Thomas S., Usmani K., Vasquez L., Vera V., Villalón D., Vinson R.,
Wall R., Wang S., Ward-Moore S., Warren R., Washington C.,
Washington S., Williams G., Williamson A., Wleczky R., Wooden S.,
Worley K., Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

and Glibbs R.
Direct Submission
Unpublished
2 (bases 1 to 150683)
Worley K.C.
Direct Submission
Submitted (30-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 10, 2000 this sequence version replaced gi:9795456.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBGN
Center clone name: RP11-683J13
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 125821 bases at least Q40
Consensus quality: 138453 bases at least Q30
Consensus quality: 144027 bases at least Q20
Estimated insert size: 142714; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 24905: contig of 24905 bp in length
* 24906 25005: gap of unknown length
* 25006 38744: contig of 13739 bp in length
* 38745 38844: gap of unknown length
* 38845 53622: contig of 14778 bp in length
* 53623 53722: gap of unknown length
* 53723 65642: contig of 11920 bp in length
* 65643 65742: gap of unknown length
* 65743 73487: contig of 7744 bp in length
* 73487 73587: gap of unknown length
* 73587 81893: contig of 8307 bp in length
* 81894 81993: gap of unknown length
* 81994 86596: contig of 6703 bp in length
* 86597 88796: gap of unknown length
* 88797 97825: contig of 9029 bp in length
* 97826 97925: gap of unknown length
* 97926 103924: contig of 5999 bp in length
* 103925 104024: gap of unknown length
* 104025 109812: contig of 5888 bp in length
* 109813 110012: gap of unknown length
* 110013 116598: contig of 6586 bp in length
* 116599 120672: gap of unknown length
* 120673 120772: contig of 3974 bp in length
* 120773 125139: gap of unknown length
* 125140 125239: contig of 4567 bp in length
* 125240 128331: gap of unknown length
* 128332 128431: contig of 3092 bp in length
* 128432 131187: gap of unknown length
* 131188 131287: contig of 2756 bp in length
* 131288 131477: gap of unknown length
* 131478 134277: contig of 2890 bp in length
* 134278 137482: gap of unknown length
* 137483 137582: contig of 3205 bp in length
* 137583 138864: gap of unknown length
* 138865 138964: contig of 1282 bp in length
* 138965 gap of unknown length

* 138965 142546: contig of 3582 bp in length
* 142547 142646: gap of unknown length
* 142647 145006: contig of 2360 bp in length
* 145007 145106: gap of unknown length
* 145107 147639: contig of 2533 bp in length
* 147640 147739: gap of unknown length
* 147740 149420: contig of 1681 bp in length
* 149421 149520: gap of unknown length
* 149521 150683: contig of 1163 bp in length.
Location/Qualifiers
1. 150683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-683J13"
BASE COUNT 40413 a 33700 c 32450 g 41878 t 2242 others
ORIGIN
Query Match 0.8%; Score 20; DB 74; Length 150683;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1588 gaccaggtgacaggtcct 1607
|||||
Db 878 GACCAGGTGACAGGTCT 859
RESULT 38
AC025326 165490 bp DNA HTG 16-JUN-2000
LOCUS Homo sapiens chromosome 2 clone RP11-621L13 map 2, WORKING DRAFT
DEFINITION
AC025326
SEQUENCE 29 unordered pieces.
AC025326.3 GI:8569175
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 165490)
Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 2, clone RP11-621L13
Unpublished
2 (bases 1 to 165490)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,N., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hages,B., Heatford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lakoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeckers,B.,
Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Plerrie,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testave,S., Theodore,J., Titrill,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wymann,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jun 16, 2000 this sequence version replaced gi:8099801.
ALL repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 621_L_13
Center clone name: L5990
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 150085 bases at least Q40
Consensus quality: 158263 bases at least Q30
Consensus quality: 161160 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 162690; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1610: contig of 1610 bp in length
* 1 1611 1710: gap of 100 bp
* 1711 2825: contig of 1115 bp in length
* 2826 2925: gap of 100 bp
* 2926 3981: contig of 1055 bp in length
* 3982 4081: gap of 100 bp
* 4082 5162: contig of 1081 bp in length
* 5163 5262: gap of 100 bp
* 5263 6695: contig of 1433 bp in length
* 6696 6795: gap of 100 bp
* 6796 8908: contig of 2113 bp in length
* 8909 9008: gap of 100 bp
* 9009 10640: contig of 1632 bp in length
* 10641 10740: gap of 100 bp
* 10741 11821: contig of 1081 bp in length
* 11822 11921: gap of 100 bp
* 11922 13344: contig of 1423 bp in length
* 13345 13444: gap of 100 bp
* 13445 14731: contig of 1287 bp in length
* 14732 14831: gap of 100 bp
* 14832 17046: contig of 2215 bp in length
* 17047 17146: gap of 100 bp
* 17147 19155: contig of 2009 bp in length
* 19156 19255: gap of 100 bp
* 19256 20955: contig of 1700 bp in length
* 20956 21055: gap of 100 bp
* 21056 23798: contig of 2743 bp in length
* 23799 23898: gap of 100 bp
* 23899 25711: contig of 1813 bp in length
* 25712 25811: gap of 100 bp
* 25812 28549: contig of 2738 bp in length
* 28550 28649: gap of 100 bp
* 28650 33094: contig of 4445 bp in length
* 33095 33194: gap of 100 bp
* 33195 38141: contig of 4947 bp in length
* 38142 38241: gap of 100 bp
* 38242 44592: contig of 6351 bp in length
* 44593 44692: gap of 100 bp
* 44693 51368: contig of 6676 bp in length
* 51369 51468: gap of 100 bp
* 51469 58005: contig of 6537 bp in length
* 58006 58105: gap of 100 bp
* 58106 66252: contig of 8147 bp in length
* 66253 66352: gap of 100 bp
* 66353 77118: contig of 10766 bp in length

```

* 7719 77218: gap of 100 bp
* 77219 85575: contig of 8357 bp in length
* 85576 85675: gap of 100 bp
* 85676 96950: contig of 11275 bp in length
* 96951 97050: gap of 100 bp
* 97051 106803: contig of 9753 bp in length
* 106804 106903: gap of 100 bp
* 106904 118011: contig of 11108 bp in length
* 118012 118111: gap of 100 bp
* 118112 133047: contig of 14936 bp in length
* 133048 133147: gap of 100 bp
* 133148 165490: contig of 32343 bp in length.
*
FEATURES
SOURCE
1..165490
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone_lib="RP11-621L13"
/clone_1bp="RPCT-11 Human Male BAC"
1..1610
misc_feature
/feature="assembly_fragment"
1711..2825
misc_feature
/feature="assembly_fragment"
2926..3981
misc_feature
/feature="assembly_fragment"
4082..5162
misc_feature
/feature="assembly_fragment"
5263..6695
misc_feature
/feature="assembly_fragment"
6796..8908
misc_feature
/feature="assembly_fragment"
9009..10640
misc_feature
/feature="assembly_fragment"
10741..11821
misc_feature
/feature="assembly_fragment"
11922..13344
misc_feature
/feature="assembly_fragment"
13445..14731
misc_feature
/feature="assembly_fragment"
14832..17046
misc_feature
/feature="assembly_fragment"
17147..19155
misc_feature
/feature="assembly_fragment"
19256..20955
misc_feature
/feature="assembly_fragment"
21056..23798
misc_feature
/feature="assembly_fragment"
23899..25711
misc_feature
/feature="assembly_fragment"
clone_end:T7
vector_side:right"
25812..28549
misc_feature
/feature="assembly_fragment"
28550..33094
misc_feature
/feature="assembly_fragment"
33195..38141
misc_feature
/feature="assembly_fragment"
38242..44592
misc_feature
/feature="assembly_fragment"
44693..51368
misc_feature
/feature="assembly_fragment"
51469..58005
misc_feature
/feature="assembly_fragment"
clone_end:SP6
vector_side:right"
58106..66252
misc_feature
/feature="assembly_fragment"
66353..77118
misc_feature
/feature="assembly_fragment"
77219..85575
misc_feature
/feature="assembly_fragment"
85676..96950

```

```

misc_feature
/feature="assembly_fragment"
97051..106803
misc_feature
/feature="assembly_fragment"
106904..118011
misc_feature
/feature="assembly_fragment"
118112..133047
misc_feature
/feature="assembly_fragment"
133148..165490
BASE COUNT 41404 a 40156 c 40246 g 40863 t 2821 others
ORIGIN
Query Match 0.8%; Score 20: DB 69; Length 165490;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2222 aagaatggagtgccctt 2241
Db 163692 AGAAGTGGAGTTCCTT 163711
|||||
RESULT 39
AC018876 165493 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 18 clone RP11-289E15, WORKING DRAFT
DEFINITION SEQUENCE, 8 unordered pieces.
ACCESSION AC018876
VERSION AC018876.4 GI:7230959
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 165493)
WATERSTON,R.H.
JOURNAL The sequence of Homo sapiens clone
2 (bases 1 to 165493)
WATERSTON,R.H.
REFERENCE Direct Submission
AUTHORS Submitted (21-DEC-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Mar 13, 2000 this sequence version replaced gi:6850795.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0289E15
----- Summary Statistics -----
Sequencing vector: M13; 81%
Sequencing vector: Plasmid; 19%
Chemistry: Dye-terminator Big Dye; 19% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 157046 bases at least Q40
Consensus quality: 160223 bases at least Q30
Consensus quality: 161942 bases at least Q20
Insert size: 174000; agarose-ff
Insert size: 164793; sum-of-contigs
Quality coverage: 4.36 in Q20 bases; agarose-ff
Quality coverage: 4.61 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 1322: contig of 1322 bp in length
* 1323 1422: gap of unknown length
* 1423 8681: contig of 7259 bp in length
* 8682 28642: contig of 19861 bp in length
* 28643 28742: gap of unknown length
* 28743 46484: contig of 17742 bp in length
* 46485 68690: contig of 22106 bp in length
* 68691 99194: contig of 30404 bp in length
* 99195 99294: gap of unknown length
* 99295 131431: contig of 32137 bp in length
* 131432 131532: gap of unknown length
* 131532 165493: contig of 33962 bp in length.
Location/Qualifiers
1. 165493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/clone="RP11-289E15"

BASE COUNT 52008 a 30658 c 30763 g 51358 t 706 others

ORIGIN

Query Match 0.8%; Score 20; DB 65; Length 165493;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1716 gccacaccccttgctgtg 1735
|||||
Db 11874 GCTTACCTTGTGCTG 11855

RESULT 40
LOCUS AL441923 165914 bp DNA HTG 07-APR-2001
DEFINITION Homo sapiens chromosome X clone RP11-104D21 map p11.21-11.23, ***
ACCESSION AL441923
VERSION AL441923.8 GI:13561280
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 165914)
Brown, A.
Direct Submission
Submitted (06-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 8, 2001 this sequence version replaced g1:13396654.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA104D21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 95% of reads
Chemistry: Dye-terminator ET-amersham; 4% of reads
Consensus quality: 164402 bases at least Q40
Consensus quality: 164723 bases at least Q30
Consensus quality: 164922 bases at least Q20
Insert size: 156718; sum-of-contigs
Insert size: 156718; 6.1% error; agarose-fp
Quality coverage: 6.95x in Q20 bases; sum-of-contigs Quality
coverage: 9.73x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 157250: contig of 157250 bp in length
* 157251 157350: gap of 100 bp
* 157351 159384: contig of 2034 bp in length
* 159385 159484: gap of 100 bp
* 159485 161555: contig of 2071 bp in length
* 161556 161655: gap of 100 bp
* 161656 163744: contig of 2089 bp in length
* 163745 163844: gap of 100 bp
* 163845 165914: contig of 2070 bp in length.
Location/Qualifiers
1. 165914
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="p11.21-11.23"
/clone="RP11-104D21"
/clone-lib="RP11-11.1"
1. 157250
/note="assembly-fragment:00834"
157351. 159384
/note="assembly-fragment:01928"
159485. 161555
/note="assembly-fragment:02600"
161656. 163744
/note="assembly-fragment:02933"
163845. 165914
/note="assembly-fragment:03272"

BASE COUNT 49534 a 32899 c 33616 g 49465 t 400 others

ORIGIN

Query Match 0.8%; Score 20; DB 81; Length 165914;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2247 catgaagctcgtctggag 2266
|||||
Db 137911 CATGAAGCTCTGCTTGAG 137892

RESULT 41
LOCUS AL162577 166344 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 6 clone RP11-238B16, *** SEQUENCING IN
PROGRESS ***; 8 unordered pieces.
ACCESSION AL162577
VERSION AL162577.11 GI:9988271
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 166344)
Sims, S.
Direct Submission
Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 6, 2000 this sequence version replaced g1:9884780.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information

Center project name: BA238B16
 ----- Summary Statistics -----
 Sequencing program: XGAP4; version 4.5
 Sequencing vector: plasmid; 108752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 162011 bases at least Q40
 Consensus quality: 163747 bases at least Q30
 Consensus quality: 164823 bases at least Q20
 Insert size: 165644; sum-of-coverage
 Insert size: 171988; 10.5% error; agarose-fp
 Quality coverage: 4.9% in Q20 bases; sum-of-coverage
 coverage: 4.9% in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved
 *
 * 1 11500: contig of 11500 bp in length
 * 11501 11600: gap of 100 bp
 * 11601 35909: contig of 24309 bp in length
 * 35910 36009: gap of 100 bp
 * 36010 39789: contig of 3780 bp in length
 * 39790 39889: gap of 100 bp
 * 39890 47363: contig of 7474 bp in length
 * 47364 47463: gap of 100 bp
 * 47464 98169: contig of 50706 bp in length
 * 98170 98269: gap of 100 bp
 * 98270 110390: contig of 11212 bp in length
 * 110391 110490: gap of 100 bp
 * 110491 122457: contig of 11967 bp in length
 * 122458 122557: gap of 100 bp
 * 122558 166344: contig of 43787 bp in length.
 *
 * Location/Qualifiers
 * 1. 166344
 * /organism="Homo sapiens"
 * /db_xref="taxon:9606"
 * /chromosome="6"
 * /clone="RP11-238B16"
 * /clone_id="RPC1-11.1"
 * 1. 11500
 * /note="assembly_fragment:00621
 * clone_end:SP6
 * vector_side:left"
 * 11601. 35909
 * /note="assembly_fragment:01398
 * fragment_chain:1"
 * 36010. 39789
 * /note="assembly_fragment:02316
 * fragment_chain:1"
 * 39890. 47363
 * /note="assembly_fragment:00526"
 * fragment_chain:1"
 * 47464. 98169
 * /note="assembly_fragment:00873"
 * fragment_chain:1"
 * 98270. 110390
 * /note="assembly_fragment:00897"
 * fragment_chain:1"
 * 110491. 122457
 * /note="assembly_fragment:01455"
 * fragment_chain:1"
 * 122558. 166344
 * /note="assembly_fragment:01966
 * fragment_chain:1"
 * clone_end:T7
 * vector_side:right"
 *
 * BASE COUNT 50249 a 32106 c 31398 g 51875 t 716 others
 * ORIGIN
 * Query Match 0.8%; Score 20; DB 79; Length 166344;
 * Best Local Similarity 100.0%; Pred. No. 25;
 * Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2272 ccacaatgcccaactgat 2291
 Db 66987 CCACCAATGCCCAACTGAT 66988
 RESULT 42
 AC021927
 LOCUS
 DEFINITION
 Homo sapiens clone RP11-29P22, WORKING DRAFT SEQUENCE, 26 unordered
 pieces.
 AC021927 168997 bp DNA HTG 03-MAR-2000
 AC021927 GI:7144973
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 168997)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Dearlilano,K., Dewar,K., Domino,M., Doyle,M., Fennestor,J.,
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
 MacDonald,P., Margulis,N., McEwan,P., McGuck,A., McKernan,K.,
 McPheters,R., Meldrum,J., Menes,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
 Pletre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tittel,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 3, 2000 this sequence version replaced gi:1684426.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WtBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4856
 Center clone name: 29_P_22
 ----- Summary Statistics -----
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 156612 bases at least Q40
 Consensus quality: 162417 bases at least Q30
 Consensus quality: 164820 bases at least Q20
 Insert size: 190000; agarose-fp
 Insert size: 166497; sum-of-coverage
 Quality coverage: 3.6 in Q20 bases; agarose-fp
 Quality coverage: 4.2 in Q20 bases; sum-of-coverage

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 26 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will

* be preserved.	1046:	contig of 1046 bp	in length
* 1	1047	1146: gap of 100 bp	
* *	1147	2513: contig of 1367 bp	in length
* *	2514	2613: gap of 100 bp	
* *	2614	3854: contig of 1241 bp	in length
* *	3855	3954: gap of 100 bp	
* *	3955	5409: contig of 1455 bp	in length
* *	5410	5509: gap of 100 bp	
* *	5510	8344: contig of 2835 bp	in length
* *	8345	8444: gap of 100 bp	
* *	8445	11304: contig of 2860 bp	in length
* *	11305	11404: gap of 100 bp	
* *	11405	14151: contig of 2747 bp	in length
* *	14152	14251: gap of 100 bp	
* *	14252	17174: contig of 2923 bp	in length
* *	17175	17274: gap of 100 bp	
* *	17275	19939: contig of 2665 bp	in length
* *	19940	20039: gap of 100 bp	
* *	20040	23411: contig of 3372 bp	in length
* *	23412	23511: gap of 100 bp	
* *	23512	27130: contig of 3619 bp	in length
* *	27131	27230: gap of 100 bp	
* *	27231	29770: contig of 2540 bp	in length
* *	29771	29870: gap of 100 bp	
* *	29871	36121: contig of 6251 bp	in length
* *	36122	36221: gap of 100 bp	
* *	36222	41978: contig of 5757 bp	in length
* *	41979	42078: gap of 100 bp	
* *	42079	49159: contig of 7081 bp	in length
* *	49160	49259: gap of 100 bp	
* *	49260	57801: contig of 8542 bp	in length
* *	57802	57901: gap of 100 bp	
* *	57902	64360: contig of 6459 bp	in length
* *	64361	64460: gap of 100 bp	
* *	64461	72524: contig of 8064 bp	in length
* *	72525	72624: gap of 100 bp	
* *	72625	81728: contig of 9104 bp	in length
* *	81729	81828: gap of 100 bp	
* *	81829	92259: contig of 10431 bp	in length
* *	92260	92359: gap of 100 bp	
* *	92360	100902: contig of 8543 bp	in length
* *	100903	101002: gap of 100 bp	
* *	101003	110082: contig of 9080 bp	in length
* *	110083	110182: gap of 100 bp	
* *	110183	124956: contig of 14774 bp	in length
* *	124957	125056: gap of 100 bp	
* *	125057	139552: contig of 14496 bp	in length
* *	139553	139652: gap of 100 bp	
* *	139653	153065: contig of 14131 bp	in length
* *	153066	153165: gap of 100 bp	
* *	153166	168997: contig of 15832 bp	in length.
FEATURES			
SOURCE			
	1.	168997	Location/Qualifiers
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="RP11-29P22"	
		/clone_lib="RPIC1-11 Human Male BAC"	
misc_feature		1..1046	
		/note="assembly_fragment"	
misc_feature		1147..2513	
		/note="assembly_fragment"	
misc_feature		2614..3854	
		/note="assembly_fragment"	
misc_feature		3955..5409	
		/note="assembly_fragment"	
misc_feature		5510..8344	
		/note="assembly_fragment"	
misc_feature		8445..11304	
		/note="assembly_fragment"	
misc_feature		11405..14151	
		/note="assembly_fragment"	
misc_feature		14252..17174	

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
LOCUS AC025191	Homo sapiens chromosome 4 clone RP11-285A15 map 4, WORKING DRAFT SEQUENCE, 43 unordered pieces.	AC025191	AC025191.2	GI:7684444	HTG: HTGS_PHASE1; HTGS_DRAFT.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	Bitnen,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,D., Barre,N., Bastien,V., Beda,F.,
Query Match	Best Local Similarity	0.8%	Score:20;	DB:66;	Length:168997;			
Matches	20; Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1443	agaatcatctctctctgga	1462					
DB	110794	AGAAATCATCTTCTCTTGGA	110813					
BASE COUNT	48830 a	34497 c	33807 g	49359 t	2504	others		
ORIGIN								
misc_feature	17275..19939	/note="assembly_fragment"						
misc_feature	20040..23411	/note="assembly_fragment"						
misc_feature	23512..27130	/note="assembly_fragment"						
misc_feature	27231..29770	/note="assembly_fragment"						
misc_feature	29871..36121	/note="assembly_fragment"						
misc_feature	36222..41978	/note="assembly_fragment"						
misc_feature	42079..49159	/note="assembly_fragment"						
misc_feature	49260..57801	/note="assembly_fragment"						
misc_feature	57902..64360	/note="assembly_fragment"						
misc_feature	64461..72524	/note="assembly_fragment"						
misc_feature	72625..81728	/note="assembly_fragment"						
misc_feature	81829..92259	/note="assembly_fragment"						
misc_feature	92360..100907	/note="assembly_fragment"						
misc_feature	clone_end:SP6							
misc_feature	vector_side:right							
misc_feature	101003..110082	/note="assembly_fragment"						
misc_feature	110183..124956	/note="assembly_fragment"						
misc_feature	125057..139532	/note="assembly_fragment"						
misc_feature	139653..153065	/note="assembly_fragment"						
misc_feature	153166..168997	/note="assembly_fragment"						

Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hages, B., Heaford, A., Horton, L.,
Hewland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Kartas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheters, R.,
Medlitz, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisanu, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tasfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, D.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 3, 2000 this sequence version replaced g1:718869.

All repeats were identified using RepeatMasker:

Smitt, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L5067

Center clone name: 285_A_15

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 150228 bases at least Q40

Consensus quality: 163346 bases at least Q30

Consensus quality: 169409 bases at least Q20

Insert size: 189000; agarose-fp

Insert size: 173668; sum-of-contigs

Quality coverage: 2.9 in Q20 bases; agarose-fp

Quality coverage: 3.2 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 43 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1215: contig of 1215 bp in length
1216 1315: gap of 100 bp
1316 2523: contig of 1208 bp in length
2524 2623: gap of 100 bp
2624 3811: contig of 1188 bp in length
3812 3911: gap of 100 bp
3912 5309: contig of 1398 bp in length
5310 5409: gap of 100 bp
5410 7250: contig of 1841 bp in length
7251 7350: gap of 100 bp
7351 8970: contig of 1620 bp in length
8971 9070: gap of 100 bp
9071 10289: contig of 1219 bp in length
10290 10389: gap of 100 bp
10390 11577: contig of 1188 bp in length
11578 11677: gap of 100 bp
11678 12960: contig of 1283 bp in length
12961 13060: gap of 100 bp
13061 14891: contig of 1831 bp in length

14892 14991: gap of 100 bp
14992 16528: contig of 1537 bp in length
16529 16628: gap of 100 bp
16629 18711: contig of 2083 bp in length
18712 18811: gap of 100 bp
18812 20214: contig of 1403 bp in length
20215 20314: gap of 100 bp
20315 22282: contig of 1968 bp in length
22283 22382: gap of 100 bp
22383 24680: contig of 2298 bp in length
24681 24780: gap of 100 bp
24781 27243: contig of 2463 bp in length
27244 27343: gap of 100 bp
27344 29231: contig of 1888 bp in length
29232 29331: gap of 100 bp
29332 32345: contig of 3014 bp in length
32346 32445: gap of 100 bp
32446 35005: contig of 2560 bp in length
35006 35105: gap of 100 bp
35106 37773: contig of 2668 bp in length
37774 37873: gap of 100 bp
37874 40633: contig of 2760 bp in length
40634 40733: gap of 100 bp
40734 43133: contig of 2400 bp in length
43134 43233: gap of 100 bp
43234 45935: contig of 2702 bp in length
45936 46035: gap of 100 bp
46036 48895: contig of 2860 bp in length
48896 48995: gap of 100 bp
48996 51307: contig of 2112 bp in length
51308 51407: gap of 100 bp
51408 55313: contig of 3906 bp in length
55314 55413: gap of 100 bp
55414 58208: contig of 2795 bp in length
58209 58308: gap of 100 bp
58309 62616: contig of 4308 bp in length
62617 62716: gap of 100 bp
62717 67413: contig of 4697 bp in length
67414 67513: gap of 100 bp
67514 70733: contig of 3220 bp in length
70734 70833: gap of 100 bp
70834 75405: contig of 4572 bp in length
75406 75505: gap of 100 bp
75506 80523: contig of 5018 bp in length
80524 80623: gap of 100 bp
80624 86066: contig of 5443 bp in length
86067 86166: gap of 100 bp
86167 92000: contig of 5834 bp in length
92001 92100: gap of 100 bp
92101 96663: contig of 4563 bp in length
96664 96763: gap of 100 bp
96764 102888: contig of 6125 bp in length
102889 102988: gap of 100 bp
102989 110032: contig of 7044 bp in length
110033 110132: gap of 100 bp
110133 117903: contig of 7771 bp in length
117904 118003: gap of 100 bp
118004 125508: contig of 7505 bp in length
125509 125608: gap of 100 bp
125609 137315: contig of 11707 bp in length
137316 137415: gap of 100 bp
137416 148418: contig of 11003 bp in length
148419 148518: gap of 100 bp
148519 161620: contig of 13102 bp in length
161621 161720: gap of 100 bp
161721 177868: contig of 16148 bp in length.

FEATURES

SOURCE

1. 177868
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-285A15"

```
misc_feature /clone_lib="RPC1-11 Human Male BAC"
1. .1215
misc_feature /note="assembly_fragment"
1316. .2523
misc_feature /note="assembly_fragment"
2624. .3811
misc_feature /note="assembly_fragment"
3912. .5309
misc_feature /note="assembly_fragment"
5410. .7250
misc_feature /note="assembly_fragment"
7351. .8970
misc_feature /note="assembly_fragment"
9071. .10289
misc_feature /note="assembly_fragment"
10390. .11577
misc_feature /note="assembly_fragment"
11678. .12960
misc_feature /note="assembly_fragment"
13061. .14891
misc_feature /note="assembly_fragment"
14992. .16528
misc_feature /note="assembly_fragment"
16629. .18711
misc_feature /note="assembly_fragment"
18812. .20214
misc_feature /note="assembly_fragment"
20315. .22282
misc_feature /note="assembly_fragment"
22383. .24680
misc_feature /note="assembly_fragment"
24781. .27243
misc_feature /note="assembly_fragment"
27344. .29231
misc_feature /note="assembly_fragment"
29332. .32345
misc_feature /note="assembly_fragment"

Query Match 0.8%; Score 20; DB 69; Length 177868;
Best Local Similarity 100.00%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1731 ggtggttgcctccaccagc 1750
|||||
Db 57383 GGTGTTGCCCCACCAGC 57402

RESULT 44
AP001899/c
LOCUS AP001899 183412 bp DNA HTG 26-JUL-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-701C7 map 18q12, WORKING
DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION AP001899.3 GI:9501833
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens DNA, clone:RP11-701C7.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 183412)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 183,412 genomic DNA of 18q12
JOURNAL Published Only in Database (2000) In press
AUTHORS 2 (bases 1 to 183412)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submissio
JOURNAL Submitted (24-APR-2000) Masahiro Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitsato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
tel:81-42-778-9923, fax:81-42-778-9924)
```

```
COMMENT
On Jul 26, 2000 this sequence version replaced gi:8117550.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center Project name: HumDrafl18
Center clone name: RP11-701C7
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 180400 bases at least Q40
Consensus quality: 181718 bases at least Q30
Consensus quality: 182304 bases at least Q20
Insert size: 182712; sum-of-contrigs
Quality coverage: 10.87x in Q20 bases; sum-of-contrigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
8 contrigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contrigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 42740 contrig of 42740 bp in length
42841 79938 contrig of 37098 bp in length
80039 109211 contrig of 26107 bp in length
109312 135418 contrig of 26298 bp in length
135519 161816 contrig of 12722 bp in length
161917 174638 contrig of 5539 bp in length
174739 180277 contrig of 3035 bp in length
180378 183412 contrig of 3035 bp in length
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 42740: contrig of 42740 bp in length
* 42741 42840: gap of 100 bp
* 42841 79938: contrig of 37098 bp in length
* 79939 80038: gap of 100 bp
* 80039 109211: contrig of 26107 bp in length
* 109212 109311: gap of 100 bp
* 109312 135418: contrig of 26107 bp in length
* 135419 135518: gap of 100 bp
* 135519 161816: contrig of 26298 bp in length
* 161817 161916: gap of 100 bp
* 161917 174638: contrig of 12722 bp in length
* 174639 174738: gap of 100 bp
* 174739 180277: contrig of 5539 bp in length
* 180278 180377: gap of 100 bp
* 180378 183412: contrig of 3035 bp in length.
FEATURES
source
1. .183412
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q12"
/clone="RP11-701C7"
1. .42740
/note="assembly_fragment"
42841. .79938
/note="assembly_fragment clone_end:SP6 vector_side:right"
80039. .109211
/note="assembly_fragment clone_end:T7 vector_side:left"
109312. .135418
/note="assembly_fragment"
135519. .161816
```

```

misc_feature      /note="assembly_fragment"
                  161917..174638
misc_feature      /note="assembly_fragment"
                  174739..180277
misc_feature      /note="assembly_fragment"
                  180378..183412
BASE COUNT      49688 a 39575 c 41016 g 52433 t      700 others
ORIGIN
Query Match      0.8% Score 20; DB 83; Length 183412;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1350 gctgcagcttcccaacttc 1369
      |||
Db 121048 GCTGCAGCTTCCCAACTTCC 121029

RESULT 45
AC009591 183855 bp DNA HTG 26-MAY-2000
LOCUS Homo sapiens chromosome 4 clone RP11-335K21 map 4, WORKING DRAFT
DEFINITION AC009591
ACCESSION AC009591
VERSION AC009591.3 GI:7107764
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 183855)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 4, clone RP11-335K21
REFERENCE 2 (bases 1 to 183855)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collins,A.,
Cooke,P., Deaellano,K., Depayre,E., Devon,J., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heath,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., McDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Nioff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tefaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
JOURNAL Direct Submission
COMMENT Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced g1:6094611.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1744
Center clone name: 335.K.21
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator-amersham; 2% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158308 bases at least Q40

```

```

Consensus quality: 170972 bases at least Q30
Consensus quality: 177434 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 183355; sum-of-contigs
Quality coverage: 5.2 in Q20 bases;
Quality coverage: 5.4 in Q20 bas.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1031: contig of 1031 bp in length
* 1032 1131: gap of 100 bp
* 1132 2356: contig of 1225 bp in length
* 2357 2456: gap of 100 bp
* 2457 3536: contig of 1080 bp in length
* 3537 3636: gap of 100 bp
* 3637 8344: contig of 4708 bp in length
* 8345 8444: gap of 100 bp
* 8445 45297: contig of 36853 bp in length
* 45298 45397: gap of 100 bp
* 45398 183855: contig of 138458 bp in length.
FEATURES
SOURCE
1. 183855
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-335K21"
/clone_id="RPCT-11 Human Male BAC"
1. 1031
/note="assembly_fragment"
1132..2356
/note="assembly_fragment"
2457..3536
/note="assembly_fragment"
3637..8344
/note="assembly_fragment"
8445..45297
/note="assembly_fragment"
clone_end="77
vector_side="right"
45398..183855
/note="assembly_fragment"
clone_end="SP6
vector_side="left"
BASE COUNT 62884 a 32653 c 31457 g 56345 t 516 others
ORIGIN
Query Match      0.8% Score 20; DB 61; Length 183855;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1731 ggtggtgcccccaaccagc 1750
      |||
Db 61148 GGTGTTGCTCCCAACACGAGC 61167

```

Search completed: October 28, 2001, 21:57:56
Job time: 12135 sec

[illegible]

	RESULT	
AAA52810	2	
ID	AAA52810	standard; cDNA; 2478 BP.

AC	AAA52810;	
XX		
DT	20-SEP-2000	(first entry)
XX		
DE	Human sulphatase G cDNA.	
XX		
KW	Human sulphatase G; hsg; chromosome 17p11.2; gene therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
FH		
FT	Key	Location/Qualifiers
FT	CDS	1..2478
FT		/*tag- a
FT		/partial
FT		/product= "hsg"
XX		
PN	W0200034327-A1.	
XX		
PD	15-JUN-2000.	
XX		
PF	09-DEC-1999;	99WO-AU01092.
XX		
PR	09-DEC-1998;	98AU-0007624.

XX	(WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.
PA	Hopwood JI, Litjens T, Hu RJ;
XX	WPI: 2000-431273/37.
DR	P-PSDB, AAY99850.
XX	
PT	Novel isolated DNA sequence which encodes human sulfatase G or its
PR	fragment useful in gene therapy for treating patients suffering from
PT	sulfatase deficiency -
XX	
PS	Claim 2; Page 29-30; 33pp; English.
XX	
CC	The present sequence encodes human sulphatase G (hSG). hSG is
CC	not a member of the well-characterised CYP6 sulfatase family.
CC	It belongs to a family showing sequence similarity to a sulphatase
CC	from the marine bacterium Alteromonas carrageenovora. The hSG gene
CC	contains 23 exons and is located at chromosome 17p11.2. The present
CC	sequence is clone lambda29.1 from a human testes cDNA library. It was
CC	isolated using human EST sequences with sequence similarity to the
CC	non-CYP6R family as a probe to screen the library. The cDNA insert was
CC	subcloned and the DNA sequence of both strands was determined. The
CC	sequence may be used to treat a patient suffering from hSG deficiency
CC	by replacing, repairing, or compensating for a DNA sequence within that
CC	patient's genome.
XX	
SQ	Sequence 2478 BP; 587 A; 686 C; 709 G; 496 T; 0 other:
Query Match	91.7%; Score 2274; DB 21; Length 2478;
Best Local Similarity	99.8%; Pred. No. 0;
Matches 2474; Conservative	0; Mismatches 4; Indels 0; Gaps 0;
YY	1 atgttgccgcttgcgtcgtcgttcgtccggcgccgaaccattgcaggaagac 60
Db	1 atgttgccgcttgcgtcgtcgttcgtccggcgccgaaccattgcaggaagac 60
YY	61 accatatcgcaaggcacccgcccgccgagcgcgccgcaagaagaccgctgcgacctg 120
Db	61 accatatcgcaaggcacccgcccgccgagcgcgccgcaagaagaccgctgcgacctg 120
YY	121 cgacgcgagagaagcgcgcgacgcgtcggggtgctccggcgcccaaacgcgttacctg 180
Db	121 cgacgcgagagaagcgcgcgacgcgtcggggtgctccggcgcccaaacgcgttacctg 180
YY	181 caggtggtgtagcagcggtgtagccggagactggcgccgcgcgttaagtcttcgcgaattc 240
Db	181 caggtggtgtagcagcggtgtagccggagactggcgccgcgcgttaagtcttcgcgaattc 240
YY	241 aacgcgtatcttccaactgtgagaagcgcttcaaagaactcatgagaagacaagtta 300
Db	241 aacgcgtatcttccaactgtgagaagcgcttcaaagaactcatgagaagacaagtta 300
YY	301 aaggtgtcgcgcctggacaacatatctctgacacgaatgacctggtctaagtltgggggc 360
Db	301 aaggtgtcgcgcctggacaacatatctctgacacgaatgacctggtctaagtltgggggc 360
YY	361 ttaagtgtgaatgatcttactctttaaaggaaacccggcttccaaagtgttactttctgta 420
Db	361 ttaagtgtgaatgatcttactctttaaaggaaacccggcttccaaagtgttactttctgta 420
YY	421 ccctcccaactgtgaaaaatafaccctggaagaacatcaaaaattttctgtgcatatgaaagga 480
Db	421 ccctcccaactgtgaaaaatafaccctggaagaacatcaaaaattttctgtgcatatgaaagga 480
YY	481 atagaactgtcgttggtggcccactctgcccagaatagcagaatgaaacatgacagt 540
Db	481 atagaactgtcgttggtggcccactctgcccagaatagcagaatgaaacatgacagt 540
YY	541 taacagatccccatatacagttgacaagagtgaggaagacccaacatggtgagagttcca 600
Db	541 taacagatccccatatacagttgacaagagtgaggaagacccaacatggtgagagttcca 600

QY 601 gaaaggctctcagcaggtctcagtcagatcttcagactccgagtcgaatgaaat 660
|||||
Db 602 gaaaggctctcagcaggtctcagtcagatcttcagactccgagtcgaatgaaat 660
QY 661 gaacacaccttcacatagtggttaagccagagaagagggttcaggactcttcgtgc 720
|||||
Db 661 gaacacaccttcacatagtggttaagccagagaagagggttcaggactcttcgtgc 720
QY 721 gtagcttcatctgttaagcttcacatlaaagaaggaactcttcgtgtcacaagcaag 780
|||||
Db 721 gtagcttcatctgttaagcttcacatlaaagaaggaactcttcgtgtcacaagcaag 780
QY 781 gagaatggcctcccaagtctgggacagctgcacatcgctcccaatcctgtcgtcacaag 840
|||||
Db 781 gagaatggcctcccaagtctgggacagctgcacatcgctcccaatcctgtcgtcacaag 840
QY 841 gggaaaagacatcatatgaagaagaagagattttgctcgaagaagctgtgtactctcca 900
|||||
Db 841 gggaaaagacatcatatgaagaagaagagattttgctcgaagaagctgtgtactctcca 900
QY 901 gatccctgctgtgctcttctgtgtgtgtagaatgtccagatgaagaagcttcacccatc 960
|||||
Db 901 gatccctgctgtgctcttctgtgtgtgtagaatgtccagatgaagaagcttcacccatc 960
QY 961 tgtgagaatgcacacttccagaggtacccaagaaagcagatgccccgtgtgctgtgtg 1020
|||||
Db 961 tgtgagaatgcacacttccagaggtacccaagaaagcagatgccccgtgtgctgtgtg 1020
QY 1021 gttcaaatggccccagcatctgtgtctgtgtgacagaggttaccaagaagtgtgagagag 1080
|||||
Db 1021 gttcaaatggccccagcatctgtgtctgtgtgacagaggttaccaagaagtgtgagagag 1080
QY 1081 ttctgggctgtgacacccagacactgtgtcctgaatgaaactgtgtcccaagcttcacaact 1140
|||||
Db 1081 ttctgggctgtgacacccagacactgtgtcctgaatgaaactgtgtcccaagcttcacaact 1140
QY 1141 cgcagcacaagaattcaaaacccagctcaactccaccccgagacatcttccccctgtcc 1200
|||||
Db 1141 cgcagcacaagaattcaaaacccagctcaactccaccccgagacatcttccccctgtcc 1200
QY 1201 accagtttccgctgtgaagaagaggccccacccctgaagtgtcccaatgttcaaggtgaa 1260
|||||
Db 1201 accagtttccgctgtgaagaagaggccccacccctgaagtgttcccaatgttcaaggtgaa 1260
QY 1261 tgcctctcctaagtlaccagatccgttcccaagagaggtgagcagaggtgccaattact 1320
|||||
Db 1261 tgcctctcctaagtlaccagatccgttcccaagagaggtgagcagaggtgccaattact 1320
QY 1321 tgcgaatcctgaggaattcaatgtgagcgctgtcagacttccaaactccagcagagcgtg 1380
|||||
Db 1321 tgcgaatcctgaggaattcaatgtgagcgctgtcagacttccaaactccagcagagcgtg 1380
QY 1381 caggaggtacaggaagagtgctgcagagagcccgaccccgacagagaagaagaagtcagtc 1440
|||||
Db 1381 caggaggtacaggaagagtgctgcagagagcccgaccccgacagagaagaagaagtcagtc 1440
QY 1441 ccagaatcatctctcttggaacaggtctgcacatcccgatgaagaatgtcagat 1500
|||||
Db 1441 ccagaatcatctctcttggaacaggtctgcacatcccgatgaagaatgtcagat 1500
QY 1501 gccacaactgttcaacataaagcccgacacgctctctgtactgtgactgtgtgtgagggcaca 1560
|||||
Db 1501 gccacaactgttcaacataaagcccgacacgctctctgtactgtgactgtgtgtgagggcaca 1560
QY 1561 ttgggacagctgtgcgcgtcatatcaggaacaggtgaggaaggttcctggcacccctgtgt 1620
|||||
Db 1561 ttgggacagctgtgcgcgtcatatcaggaacaggtgaggaaggttcctggcacccctgtgt 1620
QY 1621 gctgtgttctgttccacactgtcagcagatcaccaacacagcgttgcacaagtctctgtctg 1680
|||||
Db 1621 actgtgttctgttccacactgtcagcagatcaccaacacagcgttgcacaagtctctgtctg 1680

QY 1681 cagagagaacgcgcctgtgcatctctgggaagaagccgcttcaaccccttgctgtgtgtgcc 1740
|||||
Db 1681 cagagagaacgcgcctgtgcatctctgggaagaagccgcttcaaccccttgctgtgtgtgcc 1740
QY 1741 cccaacagcttcaaaacccctgtgtccagcagttacacaacccagtgccagaggtctctgcac 1800
|||||
Db 1741 cccaacagcttcaaaacccctgtgtccagcagttacacaacccagtgccagaggtctctgcac 1800
QY 1801 cacatcatatgattctcgtccaaatgctcttcaggaaggggtgagatcttccatctctgca 1860
|||||
Db 1801 cacatcatatgattctcgtccaaatgctcttcaggaaggggtgagatcttccatctctgca 1860
QY 1861 gtggaaagatttgatcagcttcgtgtgtgcgaacatgtgatttggaaagatttccagactgt 1920
|||||
Db 1861 gtggaaagatttgatcagcttcgtgtgtgtgcgaacatgtgatttggaaagatttccagactgt 1920
QY 1921 ctgtgtgcgcactgtcgaagcagtggttgcgtgtgtgcgtgtgtgcacacccctgtgtgaaa 1980
|||||
Db 1921 ctgtgtgcgcactgtcgaagcagtggttgcgtgtgtgcgtgtgtgcacacccctgtgtgaaa 1980
QY 1981 gtgtctatctccgggacacacatgcccctgcgaagctctgttccggagtggggaagatgcc 2040
|||||
Db 1981 gtgtctatctccgggacacacatgcccctgcgaagctctgttccggagtggggaagatgcc 2040
QY 2041 accctctcatatcagatgaagccacccctgtgaaagtgttctggaaaggaagcagttgaaaag 2100
|||||
Db 2041 accctctcatatcagatgaagccacccctgtgaaagtgttctggaaaggaagcagttgaaaag 2100
QY 2101 aacacagacacaaagtcacacccctgtgaaagtgttctggaaaggaagcagttgaaaag 2160
|||||
Db 2101 aacacagacacaaagtcacacccctgtgaaagtgttctggaaaggaagcagttgaaaag 2160
QY 2161 atgtgtaacacacttccagcagcgtatgtccaaaggtccctcttcagcccaacttcagc 2220
|||||
Db 2161 atgtgtaacacacttccagcagcgtatgtccaaaggtccctcttcagcccaacttcagc 2220
QY 2221 gagaaagtgtgagctgtgcttgcacacatgaaagtgtcgttggagacttccacaatg 2280
|||||
Db 2221 gagaaagtgtgagctgtgcttgcacacatgaaagtgtcgttggagacttccacaatg 2280
QY 2281 cccaagctgtatccccaatgaaagccctgtgtctgtgagaaatcgaaagtgtgagag 2340
|||||
Db 2281 cccaagctgtatccccaatgaaagccctgtgtctgtgagaaatcgaaagtgtgagag 2340
QY 2341 cgcagaggaagaagcggagagctgcgcaaggtgcgagcctctcttccagggagctgtgaa 2400
|||||
Db 2341 cgcagaggaagaagcggagagctgcgcaaggtgcgagcctctcttccagggagctgtgaa 2400
QY 2401 ggcggcctgtgagagatgtggagcctcagcagaagcggccacacagagagccacagcgc 2460
|||||
Db 2401 ggcggcctgtgagagatgtggagcctcagcagaagcggccacacagagagagccacagcgc 2460
QY 2461 aagaaggttcagagccag 2478
|||||
Db 2461 aagaaggttcagagccag 2478

RESULT 3
AAC76445
ID AAC76445 standard; cDNA. 2546 BP.
XX
XX AAC76445;
XX
XX 08-FEB-2001 (first entry)
DE Human ORFX ORF2000 polynucleotide sequence SEQ ID NO:3999.
XX
XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
KW vulnerability; antiportalatic; antiparkinsonian; neurotrophic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiast;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antipneumatic; antithyroid;

Db 1560 gcacacctctgctggaagtggtctatctccgagggaacacatgacctgaggtctctggt 1619
Qy 2022 ccgagtcggggaagatgcccacctctctgatacatgtaagcccacctggaagatgttggga 2081
Db 1620 ccgagtcggggaagatgcccacctctctgatacatgtaagcccacctggaagatgttggga 1679
Qy 2082 agaggaagcagctggaagaagacacacagcacacacgctcccaagccatcagcgtgggatatcg 2141
Db 1680 agaggaagcagctggaagaagacacacagcacacacgctcccaagccatcagcgtgggatatcg 1739
Qy 2142 gatgaacgcgggaattcatatgtctgaacacctctcagccacgctgatacgcaaggtccccc 2201
Db 1740 gatgaacgcgggaattcatatgtctgaacacctctcagccacgctgatacgcaaggtccccc 1799
Qy 2202 ctctcaaccccccaactcagcgaagaattggagttgctcttgaccacatgaagttcgtctt 2261
Db 1800 ctctcaaccccccaactcagcgaagaattggagttgctcttgaccacatgaagttcgtctt 1859
Qy 2262 tggagaacttccacaatgcccgaagtgtat-cccccaactgaaagccctgttgcctggcg 2320
Db 1860 tggagaacttccacaatgcccgaagtgtat-cccccaactgaaagccctgttgcctggcg 1919
Qy 2321 acatcgaaggaagatggaggaacgcgaggaaggaagcgtgcgagcgcgccgccc 2380
Db 1920 acatcgaaggaagatggaggaacgcgaggaaggaagcgtgcgagcgcgccgccc 1979
Qy 2381 tcctgtccagggagcgtgcgagcgcgctggaagatggaagcctcagcagaacgcgggcc 2440
Db 1980 tcctgtccagggagcgtgcgagcgcgctggaagatggaagcctcagcagaacgcgggcc 2039
Qy 2441 acacagaagagccacagcagccagaagaaggtcagaagccagttga 2481
Db 2040 acacagaagagccacagcagccagaagaaggtcagaagccagttga 2080

RESULT 4

AAA60207
ID AAA60207 standard; DNA; 26664 BP.

AC AAA60207;
XX

XX 07-DEC-2000 (first entry)
XX

DE Human prostate cancer predisposing gene HPC2 genomic sequence.
XX

XX Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW gene therapy; peptide therapy; drug design; ds.
XX

OS Homo sapiens.
XX

XX Homo sapiens.
XX

FX Location/Qualifiers
FT CDS
FT

FT /*tag= a
FT /product= "HPC2"
FT /note= "this sequence contains introns"
FT /transl_except= (pos:23892..23895,aa:Glu)
FT

FT /*tag= b
FT /number= 1
FT 1736..1786
FT

FT /*tag= c
FT /number= 2
FT 1925..1995
FT

FT /*tag= d
FT /number= 3
FT 3025..3089
FT

FT /*tag= e
FT /number= 4
FT 4361..4418
FT

FT /*tag= f
FT /number= 5
FT 5582..5650
FT

FT /*tag= g
FT

FT

FT

FT

FT

FT

FT

FT /number= 6
FT 7075..7194
FT /*tag= h
FT /number= 7
FT 8186..8244
FT

FT /*tag= i
FT /number= 8
FT 12878..12936
FT /*tag= j
FT /number= 9
FT 13032..13104
FT

FT /*tag= k
FT /number= 10
FT 13756..13868
FT /*tag= l
FT /number= 11
FT 15283..15378
FT

FT /*tag= m
FT /number= 12
FT 16278..16416
FT /*tag= n
FT /number= 13
FT 16498..16583
FT

FT /*tag= o
FT /number= 14
FT 18583..18701
FT /*tag= p
FT /number= 15
FT 20349..20445
FT

FT /*tag= q
FT /number= 16
FT 22172..22310
FT /*tag= r
FT /number= 17
FT 22879..22917
FT

FT /*tag= s
FT /number= 18
FT 23045..23154
FT /*tag= t
FT /number= 19
FT 23795..23895
FT

FT /*tag= u
FT /number= 20
FT 23973..24093
FT /*tag= v
FT /number= 21
FT 24354..24432
FT

FT /*tag= w
FT /number= 22
FT 25026..25170
FT /*tag= x
FT /number= 23
FT 25812..26036
FT

FT /*tag= y
FT /number= 24
FT 26447..26452
FT /*tag= z
FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

WO200027864-A1.

PD 18-MAY-2000.

PF 05-NOV-1999; 99WO-US26055.

PR 06-NOV-1998; 98US-0107468.

PA (MYRI-) MYRIAD GENETICS INC.

XX Taviglian SV, Teng DHF, Simard J, Rommens JM;

XX WPI: 2000-376481/32.

XX P-PSDB; AAB07228.

XX

XX (MYRI-) MYRIAD GENETICS INC.
XX Tavligian SV, Teng DHF, Simard J, Rommens JM;
PI WPI: 2000-376481/32.
DR P-PSDB: AAB07230.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
XX antibodies, useful for treatment and diagnosis of prostate cancer -
PS Disclosure; Page 151; 157pp; English.
XX
CC The present sequence is the coding sequence of the murine prostate
CC cancer predisposing gene Mm. HPC2, the human homologue of which is found
CC on chromosome 17p. Some alleles of this gene cause a predisposition to
CC cancer, particularly prostate cancer. This gene and its protein can be
CC used in peptide and gene therapy for cancer patients, as well as being
CC useful as diagnostic tools (both for cancer sufferers and those with a
CC predisposition to the disease) and in the production of cancer drugs.
SQ Sequence 326 BP; 39 A; 104 C; 127 G; 56 T; 0 other;

Query Match 1.0%; Score 26; DB 21; Length 326;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 166 aacaccgtgtacctgcaggtgtgtgc 191
|||||
Db 192 aacaccgtgtacctgcaggtgtgtgc 217

RESULT 7
ID AAA60373 standard; cDNA; 23 BP.
XX AAA60373;
AC
XX 07-DEC-2000 (first entry)
DT
XX Human prostate cancer predisposing gene HPC2 variant 1641insG (normal).
DE
XX Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KM gene therapy; peptide therapy; drug design; ss.
XX
XX Homo sapiens.
OS
XX WO200027864-A1.
PN
XX 18-MAY-2000.
PD
XX 05-NOV-1999; 99WO-US26055.
PF
XX 06-NOV-1998; 98US-0107468.
PR
XX (MYRI-) MYRIAD GENETICS INC.
PA
XX Tavligian SV, Teng DHF, Simard J, Rommens JM;
PI WPI: 2000-376481/32.
DR
XX
XX Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Example 5; Page 63; 157pp; English.
XX

CC The present sequence is part of the coding sequence of a variant of the
CC human prostate cancer predisposing gene HPC2, which is found on
CC chromosome 17p. This variant has been designated 1641insG. Some
CC alleles of this gene cause a predisposition to cancer, particularly
CC prostate cancer. This allele being an example of this. The HPC2 gene and
CC its protein can be used in peptide and gene therapy for cancer patients,

CC as well as being useful as diagnostic tools (both for cancer sufferers
CC and those with a predisposition to the disease) and in the production of
CC cancer drugs. This sequence was isolated by mutation screening of the
CC HPC2 gene in humans.
XX
SQ Sequence 23 BP; 5 A; 9 C; 5 G; 4 T; 0 other;

Query Match 0.9%; Score 23; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1630 gtgtccacctgcagcagatca 1652
|||||
Db 1 gtgtccacctgcagcagatca 23

RESULT 8
ID AAC24564/C standard; cDNA; 143 BP.
XX AAC24564;
AC
XX 06-OCT-2000 (first entry)
DT
XX Human secreted protein 5' EST, SEQ ID NO: 28639.
DE
XX Human secreted protein 5' EST, SEQ ID NO: 28639.
XX
KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
PN
XX 06-SEP-2000.
PD
XX 21-FEB-2000; 2000EP-0200610.
PE
XX 26-FEB-1999; 99US-0122487.
PR
XX (GEST) GENSET.
PA
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI: 2000-500381/45.
PD
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 28639; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
SQ
XX Sequence 143 BP; 40 A; 26 C; 45 G; 32 T; 0 other;

Query Match 0.8%; Score 21; DB 21; Length 143;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 976 ttccagagttaccagaagaag 996
|||||
DB 67 TTTCAGAGTACTCAGGAAG 47

RESULT 9
AAA60275/c
ID AAA60275 standard; DNA: 38 BP.

AC AAA60275;

DT 07-DEC-2000 (first entry)

DE Human HPC2 cDNA exon 1 mutation screening primer SEQ ID NO: 96.

XX Human; mouse; prostate cancer predisposing gene; HPC2;

KW human chromosome 17p; gene therapy; peptide therapy; drug design;

KW PCR primer; sequencing primer; ss.

OS Homo sapiens.

PN WO200027864-A1.

PD 18-MAY-2000.

PF 05-NOV-1999; 99WO-US26055.

PR 06-NOV-1998; 98US-0107468.

PA (MYRI-) MYRIAD GENETICS INC.

PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;

DR WPI: 2000-376481/32.

XX Human prostate cancer (HPC2) nucleic acids, polypeptides, and

PT antibodies, useful for treatment and diagnosis of prostate cancer -

XX Example 5; Page 59; 157pp; English.

XX The present sequence is a primer used in the isolation of the human

CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human

CC version of the gene is found on chromosome 17p. Some alleles cause a

CC predisposition to cancer, particularly prostate cancer. This gene and its

CC protein can be used in peptide and gene therapy for cancer patients, as

CC well as being useful as diagnostic tools (both for cancer sufferers and

CC those with a predisposition to the disease) and in the production of

CC cancer drugs.

XX Sequence 38 BP; 10 A; 9 C; 11 G; 8 T; 0 other;

XX Query Match 0.8%; Score 20; DB 21; Length 38;

XX Best Local Similarity 100.0%; Pred. No. 9.2;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 gacgcacatcatcgacga 75
|||||

DB 38 GACGCACCATATCGCAGCA 19

RESULT 10
AAF09297
ID AAF09297 standard; cDNA: 308 BP.

AC AAF09297;

XX 13-MAR-2001 (first entry)

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; Identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX Fusarium venenatum.

OS WO200056762-A2.

PN 28-SEP-2000.

PD 22-MAR-2000; 2000WO-US07781.

PF 22-MAR-1999; 99US-0273623.

PR (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

DR WPI: 2000-594572/56.

XX Monitoring differential expression of genes in filamentous fungal cells

PT uses fluorescence-labeled nucleic acids isolated from the cells and a

XX substrate of expressed sequence tags -

XX Claim 86; Page 1064; 3161pp; English.

XX The present invention describes a method for monitoring differential

CC expression of genes in a first filamentous fungal (FF) cell relative to

CC expression of the same genes in one or more second filamentous fungal

CC cells. The method uses fluorescence-labeled nucleic acids isolated from

CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs

CC are used in the methods for monitoring differential expression of genes

CC in a first filamentous fungal (FF) cell relative to expression of the

CC same genes in one or more second filamentous fungal cells. Monitoring

CC the global expression of genes from FF cells allows the production

CC potential of the microorganisms to be improved. New genes may be

CC discovered, possible functions of unknown open reading frames can be

CC identified and gene copy number variation and stability can be

CC monitored. The expression of genes can be used to study how FF cells

CC adapt to changes in culture conditions, environmental stress, spore

CC morphogenesis, recombination, metabolic or catabolic pathway

CC engineering. Using ESTs provides several advantages over genomic or

CC random cDNA clones including elimination of redundancy as one spot on an

CC array equals one gene or open reading frame, and organization of the

CC microarrays based on function of the gene products to facilitate

CC analysis of the results. AAF07478 to AAF11247 represents ESTs from

CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus

CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and

CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are

XX all specifically claimed in the present invention.

XX Sequence 308 BP; 88 A; 56 C; 100 G; 63 T; 1 other;

XX Query Match 0.8%; Score 20; DB 21; Length 308;

XX Best Local Similarity 100.0%; Pred. No. 8.7;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2334 ggagagcgacggaggaagc 2353
|||||

DB 30 ggagagcgacggaggaagc 49

RESULT 11

AAAC57035
ID AAC57035 standard; DNA: 457 BP.

AC AAC57035;

XX 25-JAN-2001 (first entry)

DE Pinus radiata transcription factor DNA sequence #481.
XX
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KM poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KM basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KM homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KM type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
XX Pinus radiata.
OS
XX WO200053724-A2.
PN
XX 14-SEP-2000.
PD
XX
XX 09-MAR-2000; 2000WO-US06112.
PF
XX 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Sheak M, Glenn M;
PI WPI; 2000-579369/54.
DR
XX
XX New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT -
XX
XX
PS Claim 1; Pages 570-571; 747pp; English.
XX
XX The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
XX
SQ Sequence 457 BP; 130 A; 94 C; 118 G; 115 T; 0 other;

Query Match 0.8%; Score 20; DB 21; Length 457;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2082 agaggaagcagtgaaaga 2101
|||||
Db 146 agaggaagcagtgaaaga 165

RESULT 12
AAA60211/C
ID AAA60211 standard; DNA; 19 BP.
XX
XX
AC AAA60211;
XX
XX 07-DEC-2000 (first entry)
XX
XX Human HPC2 CDNA 5' RACE primer SEQ ID NO: 32.
DE
XX
XX Human; mouse; prostate cancer predisposing gene; HPC2;
KM human chromosome 17p; gene therapy; peptide therapy; drug design;
KM PCR primer; sequencing primer; ss.
XX
XX Homo sapiens.
PT

XX
XX WO200027864-A1.
PN
XX 18-MAY-2000.
PD
XX 05-NOV-1999; 99WO-US26055.
PF
XX 06-NOV-1998; 98US-0107468.
PR
XX
XX (MYRI-) MYRIAD GENETICS INC.
PA
XX Tavligian SV, Teng DHF, Simard J, Rommens JM;
PI WPI; 2000-376481/32.
XX
XX
XX Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer
PT -
XX
XX
PS Example 3; Page 55; 157pp; English.
XX
XX The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
XX
SQ Sequence 19 BP; 4 A; 7 C; 3 G; 5 T; 0 other;

Query Match 0.8%; Score 19; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 257 actgtgagaagcgctca 275
|||||
Db 19 ACTGTGAGAAAGCGCTCA 1

RESULT 13
AAA60229
ID AAA60229 standard; DNA; 19 BP.
XX
XX
AC AAA60229;
XX
XX 07-DEC-2000 (first entry)
XX
XX Human HPC2 CDNA sequencing primer SEQ ID NO: 50.
DE
XX
XX Human; mouse; prostate cancer predisposing gene; HPC2;
KM human chromosome 17p; gene therapy; peptide therapy; drug design;
KM PCR primer; sequencing primer; ss.
XX
XX Homo sapiens.
OS
XX WO200027864-A1.
PN
XX 18-MAY-2000.
PD
XX 05-NOV-1999; 99WO-US26055.
PF
XX 06-NOV-1998; 98US-0107468.
PR
XX
XX (MYRI-) MYRIAD GENETICS INC.
PA
XX Tavligian SV, Teng DHF, Simard J, Rommens JM;
PI WPI; 2000-376481/32.
XX
XX Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer
PT -

Human HPC2 cDNA sequencing primer SEQ ID NO: 56.

KM Human; mouse; prostate cancer predisposing gene; HPC2;
 KW human chromosome 17p; gene therapy; peptide therapy; drug design;
 XX PCR primer; sequencing primer; ss.
 OS Homo sapiens.
 XX WO200027864-A1.
 PN 18-MAY-2000.
 PD 05-NOV-1999; 99WO-US26055.
 PF 06-NOV-1998; 98US-0107468.
 PR (MYRI-) MYRIAD GENETICS INC.
 PA Tavligran SV, Teng DHF, Simard J, Rommens JM;
 PI WPI; 2000-376481/32.
 DR
 XX
 PT Human prostate cancer (HPC2) nucleic acids, polypeptides, and
 PT antibodies, useful for treatment and diagnosis of prostate cancer -
 XX
 XX Example 3; Page 56; 157pp; English.
 CC The present sequence is a primer used in the isolation of the human
 CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
 CC version of the gene is found on chromosome 17p. Some alleles cause a
 CC predisposition to cancer, particularly prostate cancer. This gene and its
 CC protein can be used in peptide and gene therapy for cancer patients, as
 CC well as being useful as diagnostic tools (both for cancer sufferers and
 CC those with a predisposition to the disease) and in the production of
 CC cancer drugs.
 CC
 SQ Sequence 19 BP; 5 A; 3 C; 6 G; 5 T; 0 other;
 Query Match 0.8%; Score 19; DB 21; Length 19;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 751 agagggaactcttggtgc 769
 Db 1 agagggaactcttggtgc 19
 RESULT 17
 ID AAA60364 standard; DNA; 19 BP.
 AC AAA60364;
 DT 07-DEC-2000 (first entry)
 DE Human HPC2 cDNA exon 24 3'UTR mutation screening primer SEQ ID NO: 185.
 XX Human; mouse; prostate cancer predisposing gene; HPC2;
 KW human chromosome 17p; gene therapy; peptide therapy; drug design;
 KW PCR primer; sequencing primer; ss.
 OS Homo sapiens.
 XX WO200027864-A1.
 PN 18-MAY-2000.
 PD 05-NOV-1999; 99WO-US26055.
 PF 06-NOV-1998; 98US-0107468.
 PR (MYRI-) MYRIAD GENETICS INC.
 PA Tavligran SV, Teng DHF, Simard J, Rommens JM;
 PI

XX
 DR WPI; 2000-376481/32.
 XX
 PT Human prostate cancer (HPC2) nucleic acids, polypeptides, and
 PT antibodies, useful for treatment and diagnosis of prostate cancer -
 XX
 XX Example 5; Page 62; 157pp; English.
 CC The present sequence is a primer used in the isolation of the human
 CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
 CC version of the gene is found on chromosome 17p. Some alleles cause a
 CC predisposition to cancer, particularly prostate cancer. This gene and its
 CC protein can be used in peptide and gene therapy for cancer patients, as
 CC well as being useful as diagnostic tools (both for cancer sufferers and
 CC those with a predisposition to the disease) and in the production of
 CC cancer drugs.
 CC
 SQ Sequence 19 BP; 7 A; 7 C; 5 G; 0 U; 0 other;
 Query Match 0.8%; Score 19; DB 21; Length 19;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2439 ccacacagagagccacag 2457
 Db 1 ccacacagagagccacag 19
 RESULT 18
 ID AAA60214/C
 ID AAA60214 standard; DNA; 29 BP.
 AC AAA60214;
 DT 07-DEC-2000 (first entry)
 DE Human HPC2 cDNA expression construct primer SEQ ID NO: 35.
 XX
 DE Human; mouse; prostate cancer predisposing gene; HPC2;
 KW human chromosome 17p; gene therapy; peptide therapy; drug design;
 KW PCR primer; sequencing primer; ss.
 OS Homo sapiens.
 XX WO200027864-A1.
 PN 18-MAY-2000.
 PD 05-NOV-1999; 99WO-US26055.
 PF 06-NOV-1998; 98US-0107468.
 PR (MYRI-) MYRIAD GENETICS INC.
 PA Tavligran SV, Teng DHF, Simard J, Rommens JM;
 PI WPI; 2000-376481/32.
 PT Human prostate cancer (HPC2) nucleic acids, polypeptides, and
 PT antibodies, useful for treatment and diagnosis of prostate cancer -
 XX
 XX Example 3; Page 55; 157pp; English.
 CC The present sequence is a primer used in the isolation of the human
 CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
 CC version of the gene is found on chromosome 17p. Some alleles cause a
 CC predisposition to cancer, particularly prostate cancer. This gene and its
 CC protein can be used in peptide and gene therapy for cancer patients, as
 CC well as being useful as diagnostic tools (both for cancer sufferers and
 CC those with a predisposition to the disease) and in the production of
 CC cancer drugs.
 CC

SQ Sequence 29 BP; 4 A; 10 C; 7 G; 8 T; 0 other;

Query Match 0.8%; Score 19; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2463 gaagtcagagccacagta 2481
|||||
DB 29 GAAGTCAGAGCCACAGTGA 11

RESULT 19

AAA60362
ID AAA60362 standard; DNA; 37 BP.

AC AAA60362;

DT 07-DEC-2000 (first entry)

DE Human HPC2 cDNA exon 24 ORF mutation screening primer SEQ ID NO: 183.

KW Human; mouse; prostate cancer predisposing gene; HPC2;

KM human chromosome 17p; gene therapy; peptide therapy; drug design;

XX PCR primer; sequencing primer; ss.

OS Homo sapiens.

XX WO200027864-A1.

XX 18-MAY-2000.

PF 05-NOV-1999; 99WO-US26055.

PR 06-NOV-1998; 98US-0107468.

PA (MYRI-) MYRIAD GENETICS INC.

PI Tavitligian SV, Teng DHF, Simard J, Rommens JM;

DR WPI: 2000-376481/32.

PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and

PS antibodies, useful for treatment and diagnosis of prostate cancer -

XX Example 5; Page 62; 157pp; English.

CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.

XX Sequence 37 BP; 10 A; 13 C; 9 G; 5 T; 0 other;

Query Match 0.8%; Score 19; DB 21; Length 37;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2439 ccacacagagagccacag 2457
|||||
DB 19 ccacacagagagccacag 37

RESULT 20
AAA60307/C
ID AAA60307 standard; DNA; 38 BP.
XX
AC AAA60307;

XX 07-DEC-2000 (first entry)

DE Human HPC2 cDNA exon 9 mutation screening primer SEQ ID NO: 128.

KW Human; mouse; prostate cancer predisposing gene; HPC2;

KM human chromosome 17p; gene therapy; peptide therapy; drug design;

XX PCR primer; sequencing primer; ss.

OS Homo sapiens.

XX WO200027864-A1.

XX 18-MAY-2000.

PF 05-NOV-1999; 99WO-US26055.

PR 06-NOV-1998; 98US-0107468.

PA (MYRI-) MYRIAD GENETICS INC.

PI Tavitligian SV, Teng DHF, Simard J, Rommens JM;

DR WPI: 2000-376481/32.

PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and

PS antibodies, useful for treatment and diagnosis of prostate cancer -

XX Example 5; Page 60; 157pp; English.

CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.

XX Sequence 38 BP; 14 A; 6 C; 12 G; 6 T; 0 other;

Query Match 0.8%; Score 19; DB 21; Length 38;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 810 catgcctccatcattgct 828
|||||
DB 38 CATGCTCCCATCATTGCT 20

RESULT 21
AAV8751/C
ID AAV8751 standard; cDNA; 526 BP.

XX AAV8751;

XX 12-FEB-1999 (first entry)

DE EST clone HK189.

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;

KM tissue growth; actlylin; inhibin; tumour invasion suppressor; EST; human;

KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;

XX receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

XX WO9845437-A2.

XX 15-OCT-1998.

PF 10-APR-1998; 98WO-US06956.


```

PI Tavligian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI: 2000-376481/32.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
XX antibodies, useful for treatment and diagnosis of prostate cancer
XX
PS Example 3; Page 55; 157pp; English.
XX
CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Km.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 18 BP; 7 A; 5 C; 3 G; 3 T; 0 other;

Query Match 0.7%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 654 tgaaatgagccacacct 671
Db 1 tgaaatgagccacacct 18
|||||
|||||

RESULT 26
AAA60222/c
XX ID AAA60222 standard; DNA; 18 BP.
XX
XX AAA60222;
AC
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 43.
XX
KW Human; mouse; prostate cancer predisposing gene; HPC2;
RV human chromosome 17p; gene therapy; peptide therapy; drug design;
XX PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX
XX WO200027864-A1.
XX
XX 18-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US260055.
XX
XX 06-NOV-1998; 98US-0107468.
XX
XX (MYRI-) MYRIAD GENETICS INC.
XX
XX Tavligian SV, Teng DHF, Simard J, Rommens JM;
XX
XX WPI: 2000-376481/32.
XX
XX Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
XX antibodies, useful for treatment and diagnosis of prostate cancer
XX
XX Example 3; Page 55; 157pp; English.
XX
XX The present sequence is a primer used in the isolation of the human
XX and murine prostate cancer predisposing genes HPC2 and Km.HPC2. The human
XX version of the gene is found on chromosome 17p. Some alleles cause a
XX predisposition to cancer, particularly prostate cancer. This gene and its
XX protein can be used in peptide and gene therapy for cancer patients, as
XX well as being useful as diagnostic tools (both for cancer sufferers and
XX those with a predisposition to the disease) and in the production of
XX cancer drugs.
XX

```

```

XX      Sequence 18 BP; 3 A; 3 C; 5 G; 7 T; 0 other;
SQ
    Query Match          0.7%; Score 18; DB 21; Length 18;
    Best Local Similarity 100.0%; Pred. No. 89;
    Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      654 tgaatatgagccacacct 671
      |||
      18 TGAAATGAGCCACACCT 1
DB
RESULT 27
AAA60223
ID      AAA60223 standard; DNA; 18 BP.
XX
XX      AAA60223;
AC
XX      07-DEC-2000 (first entry)
DT
XX      Human HPC2 cDNA sequencing primer SEQ ID NO: 44.
DE
XX
XX      Human; mouse; prostate cancer predisposing gene: HPC2;
KM      human chromosome 17p; gene therapy; peptide therapy; drug design;
KM      PCR primer; sequencing primer; ss.
XX
XX      Homo sapiens.
OS
XX
XX      WO200027864-A1.
PN
XX      18-MAY-2000.
PD
XX
XX      05-NOV-1999; 99WO-US26055.
PF
XX      06-NOV-1998; 98US-0107468.
PR
XX      (MYRI-) MYRIAD GENETICS INC.
PA
XX      Tavligian SV, Teng DHF, Simard J, Rommens JM;
PI      WPI; 2000-376481/32.
DR
XX      Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT      antibodies, useful for treatment and diagnosis of prostate cancer -
PS      Example 3; Page 55; 157pp; English.
XX
XX      The present sequence is a primer used in the isolation of the human
CC      CC and murine prostate cancer predisposing genes HPC2 and Km.HPC2. The human
CC      version of the gene is found on chromosome 17p. Some alleles cause a
CC      predisposition to cancer, particularly prostate cancer. This gene and its
CC      protein can be used in peptide and gene therapy for cancer patients, as
CC      well as being useful as diagnostic tools (both for cancer sufferers and
CC      those with a predisposition to the disease) and in the production of
CC      cancer drugs.
CC
SQ      Sequence 18 BP; 5 A; 6 C; 2 G; 5 T; 0 other;
    Query Match          0.7%; Score 18; DB 21; Length 18;
    Best Local Similarity 100.0%; Pred. No. 89;
    Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      948 catcaaccacatctgtga 965
      |||
      1 catcaaccacatctgtga 18
DB
RESULT 28
AAA60224/C
ID      AAA60224 standard; DNA; 18 BP.
XX

```

```
AC   AAA60224;
XX   07-DEC-2000 (first entry)
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 45.
XX
XX   Human; mouse; prostate cancer predisposing gene; HPC2;
KM   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX   Homo sapiens.
OS   WO200027864-A1.
XX   18-MAY-2000.
XX   05-NOV-1999; 99WO-US26055.
XX   06-NOV-1998; 98US-0107468.
XX   (MYRI-) MYRIAD GENETICS INC.
XX   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
PI   WPI; 2000-376481/32.
XX   Human prostate cancer (HPC2) nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer
XX
XX   Example 3; Page 55; 157pp; English.
XX
XX   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Km.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
XX   Sequence 18 BP; 5 A; 2 C; 6 G; 5 T; 0 other;
SQ
Query Match          0.7%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY   948 catcaaccatctgtga 965
     |||||||||||||||
DB   18 CATCAACCATCTGTGA 1
RESULT 29
AAA60225
ID   AAA60225 standard; DNA; 18 BP.
XX
XX   AAA60225;
AC   07-DEC-2000 (first entry)
XX
XX   Human HPC2 cDNA sequencing primer SEQ ID NO: 46.
DE
XX   Human; mouse; prostate cancer predisposing gene; HPC2;
KM   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX   Homo sapiens.
OS   WO200027864-A1.
XX   18-MAY-2000.
XX   05-NOV-1999; 99WO-US26055.
```

```
XX   06-NOV-1998; 98US-0107468.
PR
XX   (MYRI-) MYRIAD GENETICS INC.
XX
XX   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
PI   WPI; 2000-376481/32.
XX
XX   Human prostate cancer (HPC2) nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer
XX
XX   Example 3; Page 55; 157pp; English.
XX
XX   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Km.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
XX   Sequence 18 BP; 5 A; 5 C; 3 G; 5 T; 0 other;
SQ
Query Match          0.7%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY   1257 tgaatgcctctcaagta 1274
     |||||||||||||||
DB   1 tgaatgcctctcaagta 18
RESULT 30
AAA60226/c
ID   AAA60226 standard; DNA; 18 BP.
XX
XX   AAA60226;
AC   07-DEC-2000 (first entry)
XX
XX   Human HPC2 cDNA sequencing primer SEQ ID NO: 47.
DE
XX   Human; mouse; prostate cancer predisposing gene; HPC2;
KM   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX   Homo sapiens.
OS   WO200027864-A1.
XX   18-MAY-2000.
XX   05-NOV-1999; 99WO-US26055.
XX   06-NOV-1998; 98US-0107468.
XX   (MYRI-) MYRIAD GENETICS INC.
XX   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
PI   WPI; 2000-376481/32.
XX   Human prostate cancer (HPC2) nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer
XX
XX   Example 3; Page 55; 157pp; English.
XX
XX   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Km.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
```

CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 18 BP; 5 A; 3 C; 5 G; 5 T; 0 other;

Query Match 0.7%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1257 tgaatgcctcctaagta 1274
|||||
DB 18 TGAATGCCCTCCTCAAGTA 1

RESULT 31
AAA60227
ID AAA60227 standard; DNA; 18 BP.
XX
AC AAA60227;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 48.
XX
XX Human; mouse; prostate cancer predisposing gene; HPC2;
KM human chromosome 17p; gene therapy; peptide therapy; drug design;
KM PCR primer; sequencing primer; ss.
XX
XX Homo sapiens.
OS
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99MO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavligian SV, Teng DHF, Sliard J, Rommens JM;
DR WPI; 2000-376481/32.
XX
PT Human prostate cancer (HPC2) nucleic acids, polypeptides, and
XX antibodies, useful for treatment and diagnosis of prostate cancer -
XX
XX Example 3; Page 55; 157pp; English.

CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 18 BP; 3 A; 3 C; 7 G; 5 T; 0 other;

Query Match 0.7%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1536 gctactgactgtgtgta 1553
|||||
DB 1 gctactgactgtgtgta 18

RESULT 32
AAA60228/c
ID AAA60228 standard; DNA; 18 BP.
XX

AC AAA60228;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 49.
XX
XX Human; mouse; prostate cancer predisposing gene; HPC2;
KM human chromosome 17p; gene therapy; peptide therapy; drug design;
KM PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99MO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavligian SV, Teng DHF, Sliard J, Rommens JM;
DR WPI; 2000-376481/32.
XX
PT Human prostate cancer (HPC2) nucleic acids, polypeptides, and
XX antibodies, useful for treatment and diagnosis of prostate cancer -
XX
XX Example 3; Page 55; 157pp; English.

CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 18 BP; 5 A; 7 C; 3 G; 3 T; 0 other;

Query Match 0.7%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1536 gctactgactgtgtgta 1553
|||||
DB 18 GCTACTGACTGTGTGTA 1

RESULT 33
AAA60232/c
ID AAA60232 standard; DNA; 18 BP.
XX

AC AAA60232;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 53.
XX
XX Human; mouse; prostate cancer predisposing gene; HPC2;
KM human chromosome 17p; gene therapy; peptide therapy; drug design;
KM PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX


```
PN WO200027864-A1.
XX
XX
PD 18-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26055.
PF
XX 06-NOV-1998; 98US-0107468.
PR
XX (MYRI-) MYRIAD GENETICS INC.
PA
XX Tavitgian SV, Teng DHF, Simard J, Rommens JM;
PI WPI; 2000-376481/32.
XX
XX
DR
XX
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
XX
PS Example 3; Page 56; 157pp; English.
XX
XX The present sequence is a primer used in the isolation of the human
XX and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
XX version of the gene is found on chromosome 17p. Some alleles cause a
XX predisposition to cancer, particularly prostate cancer. This gene and its
XX protein can be used in peptide and gene therapy for cancer patients, as
XX well as being useful as diagnostic tools (both for cancer sufferers and
XX those with a predisposition to the disease) and in the production of
XX cancer drugs.
XX
XX Sequence 18 BP; 6 A; 2 C; 7 G; 3 T; 0 other;
SQ

Query Match          0.7%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 227 tctctccgagttcaacc 244
   |||||||
Db 18 TCTTCTCCGAGTCAACC 1

RESULT 34
AAA60236
ID AAA60236 standard; DNA; 18 BP.
XX
XX AAA60236;
AC
XX
XX 07-DEC-2000 (first entry)
DT
XX
XX Human HPC2 cDNA sequencing primer SEQ ID NO: 57.
DE
XX
XX Human; mouse; prostate cancer predisposing gene; HPC2;
KW human chromosome 17p; gene therapy; peptide therapy; drug design;
KW PCR primer; sequencing primer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200027864-A1.
PN
XX
XX 18-MAY-2000.
PD
XX
XX 05-NOV-1999; 99WO-US26055.
PF
XX
XX 06-NOV-1998; 98US-0107468.
PR
XX
XX (MYRI-) MYRIAD GENETICS INC.
PA
XX
XX Tavitgian SV, Teng DHF, Simard J, Rommens JM;
PI WPI; 2000-376481/32.
XX
XX
DR
XX
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
```

```
PS Example 3; Page 56; 157pp; English.
XX
XX The present sequence is a primer used in the isolation of the human
XX and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
XX version of the gene is found on chromosome 17p. Some alleles cause a
XX predisposition to cancer, particularly prostate cancer. This gene and its
XX protein can be used in peptide and gene therapy for cancer patients, as
XX well as being useful as diagnostic tools (both for cancer sufferers and
XX those with a predisposition to the disease) and in the production of
XX cancer drugs.
XX
XX Sequence 18 BP; 8 A; 3 C; 6 G; 1 T; 0 other;
SQ

Query Match          0.7%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 986 accaaggaagcagatg 1003
   |||||||
Db 1 accaaggaagcagatg 18

RESULT 35
AAA60237
ID AAA60237 standard; DNA; 18 BP.
XX
XX AAA60237;
AC
XX
XX 07-DEC-2000 (first entry)
DT
XX
XX Human HPC2 cDNA sequencing primer SEQ ID NO: 58.
DE
XX
XX Human; mouse; prostate cancer predisposing gene; HPC2;
KW human chromosome 17p; gene therapy; peptide therapy; drug design;
KW PCR primer; sequencing primer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200027864-A1.
PN
XX
XX 18-MAY-2000.
PD
XX
XX 05-NOV-1999; 99WO-US26055.
PF
XX
XX 06-NOV-1998; 98US-0107468.
PR
XX
XX (MYRI-) MYRIAD GENETICS INC.
PA
XX
XX Tavitgian SV, Teng DHF, Simard J, Rommens JM;
PI WPI; 2000-376481/32.
XX
XX
DR
XX
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
XX
PS Example 3; Page 56; 157pp; English.
XX
XX The present sequence is a primer used in the isolation of the human
XX and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
XX version of the gene is found on chromosome 17p. Some alleles cause a
XX predisposition to cancer, particularly prostate cancer. This gene and its
XX protein can be used in peptide and gene therapy for cancer patients, as
XX well as being useful as diagnostic tools (both for cancer sufferers and
XX those with a predisposition to the disease) and in the production of
XX cancer drugs.
XX
XX Sequence 18 BP; 6 A; 7 C; 3 G; 2 T; 0 other;
SQ

Query Match          0.7%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


DR WPI; 2000-376481/32.
XX Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer
XX
XX Example 3; Page 56; 157pp; English.
XX The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
XX Sequence 18 BP; 4 A; 2 C; 7 G; 5 T; 0 other;
SQ
Query Match 0.7%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2266 gacttccacaatgcc 2283
|||||
DB 18 GACTTCCACAAATGCC 1
RESULT 39
AAZ33905
ID AAZ33905 standard; DNA; 45 BP.
XX
XX AAZ33905;
AC
XX 07-DEC-1999 (first entry)
DT
XX
DE Human PRO274 hybridisation probe.
XX
XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein; ss.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX
XX WO946281-A2.
PN
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
XX 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 22-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.

PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
XX (GETH) GENENTECH INC.
PA
XX Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
PI WPI; 1999-551358/46.
XX
XX

PI	Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
PI	Godard A, Godowski PJ, Grimaldi CJ, Gurney AJ, Hillan KJ;
PI	Kiljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
PI	Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WJ;
XX	
DR	WPI: 2000-611443/58.
XX	
PT	Novel PRO polypeptides and polynucleotides used in detection methods,
PT	to target bioactive molecules to specific cells, and to modulate
PT	cellular activities -
XX	
PS	Example 4; Page 235; 636pp; English.
XX	
CC	AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC	sequence tag) sequences which encode secreted or transmembrane PRO
CC	polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC	activity. The polynucleotides and polypeptides can be used for detecting
CC	the presence of PRO polypeptides in samples, for linking bioactive
CC	molecules to cells and for modulating biological activities of cells,
CC	using the polypeptides for specific targeting. The polypeptide targeting
CC	can be used to kill the target cells, e.g. for the treatment of cancers.
CC	The polypeptide pairs provide specific targeting of bioactive molecules
CC	to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC	the isolation of the PRO polynucleotide sequences.
XX	
SQ	Sequence 45 BP; 8 A; 10 C; 19 G; 8 T; 0 other;
XX	
Query Match	0.7%; Score 18; DB 21; Length 45;
Best Local Similarity	100.0%; Pred. No. 87;
Matches 18; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	822 catgtcgtcgtccaagga 839
Db	26 catgtcgtcgtccaagga 43
XX	
RESULT 41	
AA18066/C	
ID	AA18066 standard; DNA; 177 BP.
XX	
AC	AA18066;
XX	
DT	04-MAY-1999 (first entry)
XX	
DE	Coding sequence for human SI binding protein SN145.
XX	
KM	Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
KM	D2H; hPEP1; human; GI tract receptor; sucrose-isomaltase complex;
KM	intestinal peptide-associated transporter; hypertension; diabetes;
KW	osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KX	therapeutic agent delivery; therapy; ss.
XX	
OS	Homo sapiens.
XX	
PN	W09851325-A2.
XX	
PD	19-NOV-1998.
XX	
PF	15-MAY-1998; 98WO-US10088.
XX	
PR	15-MAY-1997; 97US-0046595.
XX	
PA	(CYTO-) CYTOGEN CORP.
PA	(ELAN-) ELAN CORP PLC.
XX	
PI	Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
PI	Omahony DJ, Patterson CA, Singleton J;
XX	
DR	WPI: 1999-009568/01.
XX	
PT	New proteins that bind specifically to receptors in the
PT	gastro-intestinal tract and related nucleic acid - chimeras and

PT antibodies, used to deliver therapeutic or diagnostic agents to, or
PT through, the gastrointestinal tract, e.g. insulin or leuprolide
XX
PS Claim 49; Page 56; 294pp; English.
XX
CC This sequence encodes a peptide that specifically binds to the human
CC sucrase-isomaltase complex. The invention relates to purified
CC proteins (I) that bind specifically to at least one of the
CC gastro-intestinal (GI) tract receptors human intestinal
CC peptide-associated transporter (HPT1), hPEPT1, D2H and human
CC sucrase-isomaltase complex (HSI). (I) provide active transport of
CC therapeutic agents through human and animal GI tissue (into the blood)
CC for in vivo delivery, particularly for treatment or prevention of
CC hyperemesis, diabetes, osteoporosis, haemophilia, anaemia, cancer,
CC migraine, or angina pectoris. Specifically they are used to deliver
CC insulin or leuprolide, but many other suitable therapeutic agents are
CC disclosed. Including genes or inhibitory nucleic acid, imaging agents and
CC antigens. (I) may also provide targeting to the GI tract. Other uses of
CC (I) are: (i) to determine the level of specified receptors in a sample
CC (in a binding assay); and (ii) to screen for molecules that bind (I).
CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC and in immunoassays. The antibodies are used to locate, detect and
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.
SO Sequence 177 BP; 31 A; 51 C; 53 G; 42 T; 0 other;

Query Match 0.7%; Score 18; DB 20; Length 177;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1766 agcagttaccacaccagt 1783
|||||
DB 90 AGCACTACCAACACAGT 73

RESULT 42
AAT67773/C
ID AAT67773 standard; DNA: 531 BP.
XX
AC AAT67773;
XX
DT 29-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein ORF 24824087.aa.
XX
KW Vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW cytoplasmic; ds.
XX
XX Helicobacter pylori.
OS
FH Key Location/Qualifiers
FT CDS 1..531
FT /tag= a
FT /transl_except= (pos: 460..462, aa: Xaa)
FT /transl_except= (pos: 520..522, aa: Xaa)
FT /transl_except= (pos: 526..528, aa: Xaa)
FT /note= "Xaa = Unknown"
XX
XX W09640893-A1.
XX
XX 19-DEC-1996.
XX
XX 06-JUN-1996; 96WO-US09122.
XX
XX 01-APR-1996; 96US-0630405.
XX 07-JUN-1995; 95US-0487032.
XX
XX (ASTR) ASTRA AB.
XX

PI Berglindh OT, Smith D, Mellgaerd BL;
XX
XX WPI: 1997-052306/05.
DR P-PSDB; AAM20335.
XX
XX Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
PS Claim 9; Page -: 1481pp; English.
XX
XX The present sequence encodes a Helicobacter pylori cytoplasmic
CC protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the
CC bacterial DNA. The sequences were analysed for ORF of at least 180
CC nucleotides, and the predicted coding regions defined by computer
CC evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences
CC of interest, particular regions can be isolated from H. pylori by
CC PCR amplification for recombinant polypeptide production, e.g. in
CC E. coli hosts.
CC Note: This DNA sequence is not reproduced in the specification and
CC has been derived from the related specification, W09719098.
SO Sequence 531 BP; 175 A; 88 C; 114 G; 151 T; 3 other;

Query Match 0.7%; Score 18; DB 18; Length 531;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1710 aaagcgcttcacccctt 1727
|||||
DB 311 AAAGCGCTTCACCTTT 294

RESULT 43
AAT77453/C
ID AAT77453 standard; DNA: 531 BP.
XX
AC AAT77453;
XX
DT 11-AUG-1997 (first entry)
XX
DE H. pylori cytoplasmic protein ORF 24824087.aa.
XX
KW Chronic gastritis; duodenal ulcer disease; activator;
KW inhibitor; bacterial life cycle; vaccine; immunisation; detection;
KW antisense; inhibition; cytoplasmic; Na+/H+ antiporter;
KW Escherichia coli; ds.
XX
XX Helicobacter pylori.
OS
FH Key Location/Qualifiers
FT CDS 1..531
FT /tag= a
FT /transl_except= (pos: 460..462, aa: Xaa)
FT /transl_except= (pos: 520..522, aa: Xaa)
FT /transl_except= (pos: 526..528, aa: Xaa)
FT /note= "Xaa = Unknown"
XX
XX W09719098-A1.
XX
XX 29-MAY-1997.
XX
XX 15-NOV-1996; 96WO-US18542.
XX

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 0.7%; Score 18; DB 21; Length 791;
Best Local Similarity 100.0%; Pred.No. 81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2067 ggaagatggttggaga 2084
|||||
Db 508 ggaagatggttggaga 525

Search completed: October 28, 2001, 21:04:24
Job time: 5363 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 19:32:56 ; Search time 66.26 Seconds

(without alignments)
7088.453 Million cell updates/sec

Title: US-09-434-382-1

Perfect score: 2481

Sequence: 1 atgtggcgcttgctgcgt.....agaagtcagaccacagta 2481

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 9465562 residues

Word size : 8

Total number of hits satisfying chosen parameters: 376375

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCBUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	0.8	2955	2	US-08-867-941-9
2	19	0.8	2955	4	US-09-074-658-9
3	19	0.8	3000	2	US-08-867-941-8
4	19	0.8	3000	4	US-09-074-658-8
5	19	0.8	7641	2	US-08-867-941-6
6	19	0.8	7641	4	US-09-074-658-6
7	18	0.7	2186	2	US-08-878-546-9
8	18	0.7	2605	2	US-08-680-395-4
9	17	0.7	289	2	US-08-967-101-23
10	17	0.7	289	2	US-08-592-541-23
11	17	0.7	288	3	US-09-124-698-23
12	17	0.7	289	4	US-09-127-480-23
13	17	0.7	289	4	US-08-496-841C-23
14	17	0.7	304	2	US-08-611-757-20
15	17	0.7	304	5	PCT-US95-05980-20
16	17	0.7	1084	2	US-08-184-009-110
17	17	0.7	1084	2	US-08-458-356-110
18	17	0.7	1094	2	US-08-184-009-109
19	17	0.7	1094	2	US-08-458-356-109
20	17	0.7	1330	2	US-09-118-442-29
21	17	0.7	1691	3	US-08-993-118-8
22	17	0.7	1691	3	US-08-845-528C-8
23	17	0.7	1816	2	US-08-951-148-2
24	17	0.7	1816	2	US-09-165-234-2
25	17	0.7	1816	3	US-09-274-570-2
26	17	0.7	2419	1	US-07-807-043B-7
27	17	0.7	2419	1	US-08-299-649B-7

28	17	0.7	2419	2	US-08-142-368A-7	Sequence 7, Appl
29	17	0.7	2419	3	US-08-967-727-7	Sequence 7, Appl
30	17	0.7	2419	4	US-08-037-230D-7	Sequence 7, Appl
31	17	0.7	2420	1	US-08-465-167A-23	Sequence 23, Appl
32	17	0.7	3457	1	US-08-295-882-1	Sequence 1, Appl
33	17	0.7	3796	1	US-08-343-760A-1	Sequence 1, Appl
34	17	0.7	4488	1	US-08-441-430-1	Sequence 1, Appl
35	17	0.7	5674	1	US-07-807-043B-8	Sequence 8, Appl
36	17	0.7	5674	1	US-08-180-411A-1	Sequence 1, Appl
37	17	0.7	5674	1	US-08-299-849B-8	Sequence 8, Appl
38	17	0.7	5674	2	US-08-560-024-1	Sequence 1, Appl
39	17	0.7	5674	2	US-08-142-368A-8	Sequence 8, Appl
40	17	0.7	5674	3	US-08-967-727-8	Sequence 8, Appl
41	17	0.7	5674	4	US-08-037-230D-8	Sequence 8, Appl
42	17	0.7	6464	1	US-08-321-478-6	Sequence 6, Appl
43	17	0.7	11236	1	US-07-853-913-7	Sequence 1, Appl
44	17	0.7	29604	3	US-08-781-891-207	Sequence 207, App
45	17	0.7	68750	3	US-09-335-409-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-867-941-9
; Sequence 9, Application US/08867941
; Patent No. 5977337
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Du, Run-Pan
; APPLICANT: Wang, Qunjun
; APPLICANT: Yang, Yan-ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & Mcburney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,941
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-681 MTS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2955 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-867-941-9

Query Match 0.8%; Score 19; DB 2; Length 2955;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 761 tcttggtctcaagcaaa 779
|||||

Db 591 TCTTGCTCAAGCAA 609

RESULT 2

US-09-074-658-9

; Sequence 9, Application US/09074658

; Patent No. 6184371

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M

; APPLICANT: Run-Pan Du

; APPLICANT: Quijun Wang

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H

; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA

; NUMBER OF SEQUENCES: 78

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 6th Floor, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/074,658

; FILING DATE: 08-MAY-1998

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-795

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2955 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-074-658-9

Query Match 0.8%; Score 19; DB 4; Length 2955;

Best Local Similarity 100.0%; Pred. No. 9.4;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 761 tcttggtctcaagcaaa 779
|||||

Db 591 TCTTGCTCAAGCAA 609

RESULT 3

US-08-867-941-8

; Sequence 8, Application US/08867941

; Patent No. 5977337

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M

; APPLICANT: Du, Run-Pan

; APPLICANT: Wang, Quijun

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H

; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA

; NUMBER OF SEQUENCES: 67

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 6th Floor, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/867,941

FILING DATE: 03-JUN-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-681 MIS:jfb

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 3000 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-867-941-8

Query Match 0.8%; Score 19; DB 2; Length 3000;

Best Local Similarity 100.0%; Pred. No. 9.4;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 761 tcttggtctcaagcaaa 779
|||||

Db 636 TCTTGCTCAAGCAA 654

RESULT 4

US-09-074-658-8

; Sequence 8, Application US/09074658

; Patent No. 6184371

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M

; APPLICANT: Run-Pan Du

; APPLICANT: Quijun Wang

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H

; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA

; NUMBER OF SEQUENCES: 78

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 6th Floor, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/074,658

FILING DATE: 08-MAY-1998

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-795

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 3000 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-658-8

Query Match 0.8%; Score 19; DB 4; Length 3000;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 tcttggtctcaagcaaa 779
|||||
DB 636 tcttggtctcaagcaaa 654

RESULT 5
US-08-867-941-6
Sequence 6, Application US/08867941
Patent No. 5977337
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Du, Run-Pan
APPLICANT: Wang, Quijun
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,941
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 7641 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-867-941-6

Query Match 0.8%; Score 19; DB 2; Length 7641;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 tcttggtctcaagcaaa 779
|||||
DB 3656 tcttggtctcaagcaaa 3674

RESULT 6
US-09-074-658-6
Sequence 6, Application US/09074658
Patent No. 6184371

GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Run-Pan Du
APPLICANT: Quijun Wang
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,658
FILING DATE: 08-MAY-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-795
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 7641 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-658-6

Query Match 0.8%; Score 19; DB 4; Length 7641;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 tcttggtctcaagcaaa 779
|||||
DB 3656 tcttggtctcaagcaaa 3674

RESULT 7
US-08-878-546-9
Sequence 9, Application US/08878546
Patent No. 5952463
GENERAL INFORMATION:
APPLICANT: SHIBANO, YUTII
APPLICANT: KIKUCHI, NORIHISA
APPLICANT: ODA, KOHEI
TITLE OF INVENTION: NOVEL PROTEINASE INHIBITOR AND
TITLE OF INVENTION: GENE ENCODING THE INHIBITOR
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: STEINBERG, RASKIN & DAVIDSON P.C.
STREET: 1140 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/878,546
FILING DATE: 19-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 158677/1996
FILING DATE: 19-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 224104/1996
FILING DATE: 26-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 48101/1997
FILING DATE: 03-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: DAVIDSON, CLIFFORD M.
REGISTRATION NUMBER: 32,728
REFERENCE/DOCKET NUMBER: 382,1009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)-768-3800
TELEFAX: (212)382-2124
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2186 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: STREPTOMYCES PLATENSIS
STRAIN: Q268
FEATURE:
NAME/KEY: CDS
LOCATION: 1477..1911
US-08-878-546-9

Query Match 0.7%; Score 18; DB 2; Length 2186;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1602 ggtccctgggcaccctgac 1619
|||||
DB 1518 ggtcctgggcaccctgac 1535

RESULT 8
US-08-680-395-4/c
Sequence 4, Application US/08680395
Patent No. 5892010
GENERAL INFORMATION:
APPLICANT: Gray, Joe W.
APPLICANT: Collins, Colin
APPLICANT: Huang, Soo-in
APPLICANT: Godfrey, Tony
APPLICANT: Kowbel, David
TITLE OF INVENTION: Genes from the 20q13 Amplicon and Their
TITLE OF INVENTION: Uses
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/680,395
FILING DATE: 15-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-068900US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2605 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: -
LOCATION: 1..2605
OTHER INFORMATION: /note="cDNA clone cc43 of 4 kb
US-08-680-395-4
OTHER INFORMATION: transcript"

Query Match 0.7%; Score 18; DB 2; Length 2605;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 398 ttccaagtgtgacttt 415
|||||
DB 1421 ttccaagtgtgacttt 1404

RESULT 9
US-08-967-101-23/c
Sequence 23, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: FRASER, PAUL E
APPLICANT: ROMMENS, JOHANNA M
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-967-101-23

Query Match 0.7%; Score 17; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2081 aagaggaagcagtgga 2097
|||||
Db 267 AAGAGGAAGCAGTGGA 251

RESULT 10
US-08-592-541-23/c
Sequence 23, Application US/08592541
Patent No. 5986054
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592.541
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592.541
FILING DATE:
MOLECULE TYPE: DNA (genomic)
US-08-592-541-23

Query Match 0.7%; Score 17; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2081 aagaggaagcagtgga 2097
|||||
Db 267 AAGAGGAAGCAGTGGA 251

RESULT 11
US-09-124-698-23/c
Sequence 23, Application US/09124698
Patent No. 6117978
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M

APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124.698
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592.541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/127.480
FILING DATE:
MOLECULE TYPE: DNA (genomic)
US-09-124-698-23

Query Match 0.7%; Score 17; DB 3; Length 289;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2081 aagaggaagcagtgga 2097
|||||
Db 267 AAGAGGAAGCAGTGGA 251

RESULT 12
US-09-127-480-23/c
Sequence 23, Application US/09127480
Patent No. 6194153
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/127.480
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-127-480-23

Query Match 0.7%; Score 17; DB 4; Length 289;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2081 aagaggaagcagtgga 2097
|||||
DB 267 AAGAGGAGCAGTGGA 251

RESULT 13
US-08-496-841C-23/C
Sequence 23, Application US/08496841C
Patent No. 6210919
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
ROMMENS, JOHANNA M
FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby, PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,841C
FILING DATE: 28-Jun-1995
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Paul F. Fehner, Ph.D.
REGISTRATION NUMBER: 35,135
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-08-496-841C-23

Query Match 0.7%; Score 17; DB 4; Length 289;

Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2081 aagaggaagcagtgga 2097
|||||
DB 267 AAGAGGAGCAGTGGA 251

RESULT 14
US-08-611-757-20/C
Sequence 20, Application US/08611757
Patent No. 5859230
GENERAL INFORMATION:
APPLICANT: Kim, Jungsh P.
APPLICANT: Reyes, Gregory R.
APPLICANT: Wages, John
APPLICANT: Zhang-Keck, Zhen-Yang
APPLICANT: Young, Lavonne
TITLE OF INVENTION: No. 5859230-A/No. 5859230-B/No. 5859230-C/No. 5859230-D/No.
TITLE OF INVENTION: Agents and Molecular Cloning Thereof
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,757
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/246,985
FILING DATE: 20-MAY-1994
APPLICATION NUMBER: US 025,396
FILING DATE: 24-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/922,493
FILING DATE: 30-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fadian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: MY 190 Clone D30
US-08-611-757-20

Query Match 0.7%; Score 17; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 cgcgcgagcgccgcg 98
|||||
DB 22 CGCCGCGAGCGCGCG 6

RESULT 15
PCT-US95-05980-20/c
; Sequence 20, Application PC/TUS9505980
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: Non-A/Non-B/Non-C/Non-D/Non-E Hepatitis
; TITLE OF INVENTION: Agents and Molecular Cloning Theroot
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05980
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,986
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 304 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: My 190 Clone D30
; PCT-US95-05980-20

Query Match 0.7%; Score 17; DB 5; Length 304;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 82 GCGCGGAGGCGCGG 98
Db 22 GCGCGGAGGCGCGG 6

RESULT 16
US-08-184-009-110
; Sequence 110, Application US/08184009
; Patent No. 5833975
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtls, Morris & Safford
; STREET: 530 Fifth Avenue

CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/184,009
; FILING DATE: 19-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1084 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-184-009-110

Query Match 0.7%; Score 17; DB 2; Length 1084;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1602 GTCCTGGGCAACCTGG 1618
Db 292 GGTCTGGGCAACCTGG 308

RESULT 17
US-08-458-356-110
; Sequence 110, Application US/08458356
; Patent No. 5942235
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtls, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,356
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,009
; FILING DATE: 19-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 1084 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-458-356-110

Query Match 0.7%; Score 17; DB 2; Length 1084;
Best local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1602 ggtcctgggacacctgg 1618
|||||

DB 292 GGTCTGGGACCCCTGG 308

RESULT 18
US-08-184-009-109
Sequence 109, Application US/08184009
Patent No. 5833975
GENERAL INFORMATION:
APPLICANT: Paolletti, Enzo
APPLICANT: Tartaglia, James
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/184,009
FILING DATE: 19-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-0712
TELEFAX: (212) 840-3333
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 1094 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-184-009-109

Query Match 0.7%; Score 17; DB 2; Length 1094;
Best local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1602 ggtcctgggacacctgg 1618
|||||

DB 314 GGTCTGGGACCCCTGG 330

RESULT 19
US-08-458-356-109
Sequence 109, Application US/08458356
Patent No. 5942235
GENERAL INFORMATION:
APPLICANT: Paolletti, Enzo
APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,356
FILING DATE: 02-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/184,009
FILING DATE: 19-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 1094 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-458-356-109

Query Match 0.7%; Score 17; DB 2; Length 1094;
Best local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1602 ggtcctgggacacctgg 1618
|||||

DB 314 GGTCTGGGACCCCTGG 330

RESULT 20
US-09-118-442-29/C
Sequence 29, Application US/09118442B
Patent No. 6197561
GENERAL INFORMATION:
APPLICANT: Martino-Catt, Susan J.
APPLICANT: Wang, Hongyu
APPLICANT: Beach, Larry R.
APPLICANT: Wang, Xun
APPLICANT: Bowen, Benjamin A.
TITLE OF INVENTION: Genes Controlling Phytate Metabolism in
Plants and Uses Thereof
FILE REFERENCE: 0706
CURRENT APPLICATION NUMBER: US/09/118,442B

CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/055,446
; EARLIER FILING DATE: 1997-08-11
; EARLIER APPLICATION NUMBER: 60/055,526
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: 60/053,944
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 1330
; TYPE: DNA
; ORGANISM: Zea mays
US-09-118-442-29

Query Match 0.7%; Score 17; DB 4; Length 1330;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1473 catccgatgaagattc 1489
Db 493 catccgatgaagattc 477

RESULT 21
US-08-993-118-8

; Sequence 8, Application US/08993118
; Patent No. 5997872
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie;
; APPLICANT: DE SMET, Charles;
; APPLICANT: BOON-FALLEER, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGC-C1 AND USES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,118
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,528
; FILING DATE: April 25, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1691 base pairs
; TYPE: nucleotides
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
US-08-993-118-8

Query Match 0.7%; Score 17; DB 2; Length 1691;

Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1602 ggtcctggcaccctgg 1618
Db 317 ggtcctggcaccctgg 333

RESULT 22

US-08-845-528C-8
; Sequence 8, Application US/08845528C
; Patent No. 6027924
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie;
; APPLICANT: DE SMET, Charles;
; APPLICANT: BOON-FALLEER, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGC-C1 AND USES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,528C
; FILING DATE: April 25, 1997
; CLASSIFICATION: 4335
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1691 base pairs
; TYPE: nucleotides
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
US-08-845-528C-8

Query Match 0.7%; Score 17; DB 3; Length 1691;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1602 ggtcctggcaccctgg 1618
Db 317 ggtcctggcaccctgg 333

RESULT 23

US-08-951-148-2
; Sequence 2, Application US/08951148
; Patent No. 5871973
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: CELL DIVISION REGULATORS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,148
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0407 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1816 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SPLNFZT01
CLONE: 26459
US-08-951-148-2

Query Match 0.7%; Score 17; DB 2; Length 1816;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 754 ggaactcttggtgc 770
|||||

Db 717 GGAACCTCTGTGTC 733

RESULT 24
US-09-165-234-2
Sequence 2, Application US/09165234
Patent No. 5928899
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Shah, Puri
TITLE OF INVENTION: CELL DIVISION REGULATORS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/165,234
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/951,148
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0407 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1816 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SPLNFZT01
CLONE: 26459
US-09-165-234-2

Query Match 0.7%; Score 17; DB 2; Length 1816;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 754 ggaactcttggtgc 770
|||||

Db 717 GGAACCTCTGTGTC 733

RESULT 25
US-09-274-570-2
Sequence 2, Application US/09274570
Patent No. 6121019
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Shah, Puri
TITLE OF INVENTION: CELL DIVISION REGULATORS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/274,570
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/951,148
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0407 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 1816 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SPLNFZT01
CLONE: 26459
US-09-274-570-2

Query Match 0.7%; Score 17; DB 3; Length 1816;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 754 ggaactcttggtcgt 770
|||||
Db 717 GGAACCTCTGTGCT 733

RESULT 26
US-07-807-043B-7
Sequence 7, Application US/07807043B
Patent No. 5342774
GENERAL INFORMATION:
APPLICANT: Boon, Thierry, Van den Eynde, Beno t
TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
NUMBER OF INVENTION: Rejection Antigens and Uses Thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/807,043B
FILING DATE: 19911212
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5342774man D.
REGISTRATION/DOCKET NUMBER: 30,946
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2419 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: singular
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-07-807-043B-7

Query Match 0.7%; Score 17; DB 1; Length 2419;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1602 ggtcctgggacacctg 1618
|||||
Db 739 GGTCTGGGACACCTGG 755

RESULT 27
US-08-299-849B-7
Sequence 7, Application US/08299849B
Patent No. 5612201
GENERAL INFORMATION:
APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;
APPLICANT: Chomez, Patrick
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
NUMBER OF SEQUENCES: 48
DETERMINING EXPRESSION OF A Tumor Antigen Precursor
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,849B
FILING DATE: 1-SEPTEMBER-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/037,230
FILING DATE: 26-MARCH-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5612201man D.
REGISTRATION/DOCKET NUMBER: 30,946
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-299-849B-7

Query Match 0.7%; Score 17; DB 1; Length 2419;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1602 ggtcctgggacacctg 1618
|||||
Db 739 GGTCTGGGACACCTGG 755

RESULT 28
US-08-142-368A-7
; Sequence 7, Application US/08142368A
; Patent No. 5925729
; GENERAL INFORMATION:
; APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
; APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
; APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felte & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/142,368A
; FILING DATE: 02-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; APPLICATION NUMBER: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5925729man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5253.4-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-142-368A-7

Query Match 0.7%; Score 17; DB 2; Length 2419;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1602 ggtcctgggacccctgg 1618
|||||
DB 739 ggtcctgggacccctgg 755

RESULT 29
US-08-967-727-7
; Sequence 7, Application US/08967727
; Patent No. 6025474

; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: Van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felte & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,727
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,365
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6025474man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-967-727-7

Query Match 0.7%; Score 17; DB 3; Length 2419;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1602 ggtcctgggacccctgg 1618
|||||
DB 739 ggtcctgggacccctgg 755

RESULT 30
US-08-037-230D-7
; Sequence 7, Application US/08037230D
; Patent No. 6235525
; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: Van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof

NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/037,230D
FILING DATE: 26-MARCH-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6235525man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5353
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-037-230D-7

Query Match 0.7%; Score 17; DB 4; Length 2419;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1602 ggtcctgggcacctgg 1618
|||||

DB 739 ggtcctgggcacctgg 755

RESULT 31
US-08-465-167A-23
Sequence 23, Application US/08465167A
Patent No. 5750395
GENERAL INFORMATION:
APPLICANT: Fikes, John D.
APPLICANT: Livingston, Brian D.
APPLICANT: Settle, Alessandro D.
APPLICANT: Sidney, John C.
TITLE OF INVENTION: DNA ENCODING MAGE-1 C-TERMINAL
TITLE OF INVENTION: IMMUNOGENIC PEPTIDES (as amended)
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco

STATE: CA
COUNTRY: USA
ZIP: 98111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,167A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/103,623
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14137-60-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2420 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: CDS
LOCATION: 626..1552
US-08-465-167A-23

Query Match 0.7%; Score 17; DB 1; Length 2420;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1602 ggtcctgggcacctgg 1618
|||||

DB 739 ggtcctgggcacctgg 755

RESULT 32
US-08-295-882-1/C
Sequence 1, Application US/08295882
Patent No. 5569833
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: METHOD FOR ENHANCING
TITLE OF INVENTION: PLANT PRECOCITY AND/OR REDUCING THE
TITLE OF INVENTION: STORED NITRATE CONTENT OF A PLANT
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie, Parker & Hale
STREET: P.O. Box 7068
CITY: Pasadena
STATE: CA
COUNTRY: USA
ZIP: 91109-7068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORD PERFECT 5.1 release
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/295,882
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00222
FILING DATE: March 5, 1993
APPLICATION NUMBER: 92 02658

FILING DATE: March 5, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20,958
REFERENCE/DOCKET NUMBER: 27209/DBP
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3457 base pairs
TYPE: nucleotide with corresponding
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE: (1.v1.A) ORGANISM: Nicotiana tabacum
ORIGINAL SOURCE: (1.v1.B) CELL LINE: N. tabacum cv. Xanthi
ORIGINAL SOURCE: XHFD 8
IMMEDIATE SOURCE: leaf
FEATURE:
NAME/KEY: Nitrate reductase
LOCATION: from 1 to 143 bp: Leader
LOCATION: non translated 5 sequence (leader)
LOCATION: from 144 to 2855 bp: coding sequence
LOCATION: for nitrate reductase apoenzyme
LOCATION: from 2856 to 3457 bp: non translated
LOCATION: 3 sequence
US-08-295-882-1

Query Match 0.7%; Score 17; DB 1; Length 3457;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Yy 859 gaaggaagagagattt 875
|||||
Db 94 GAAGGAAGAGATT 78

RESULT 33
US-08-343-760A-1
Sequence 1, Application US/08343760A
Patent No. 5679783
GENERAL INFORMATION:
APPLICANT: De Robertis, Edward M
APPLICANT: Sasai, Yoshiki
TITLE OF INVENTION: Tissue Differentiation Affecting
TITLE OF INVENTION: Factor and Composition
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Majestic, Parsons, Siebert & Hsue
STREET: Four Embarcadero Center, Suite 1450
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,760A
FILING DATE: 22-NOV-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Siebert, J. Suzanne
REGISTRATION NUMBER: 28,758
REFERENCE/DOCKET NUMBER: 3100.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 363-5556
TELEFAX: (415) 362-5418
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 3796 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-343-760A-1

Query Match 0.7%; Score 17; DB 1; Length 3796;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Yy 220 ctctacgtctctccga 236
|||||
Db 1036 CTCACGCTCTCCGA 1052

RESULT 34
US-08-441-430-1
Sequence 1, Application US/08441430
Patent No. 5681942
GENERAL INFORMATION:
APPLICANT: Buchwald, Manuel
APPLICANT: Strathdee, Craig A.
APPLICANT: Wevrick, Rachel
APPLICANT: Mathew, Christopher George Porter
TITLE OF INVENTION: Fanconi Anemia Type C Gene
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Polley, Esq.
ADDRESSEE: Klarquist, Sparkman, Campbell, Leigh &
ADDRESSEE: Winston, LLP
STREET: 121 S.W. Salmon, Suite 1600
CITY: Portland
STATE: Oregon
COUNTRY: U.S.A.
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: Disk, 3+/-inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS DOS
SOFTWARE: Wordperfect 5.1/ASCII Text File
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,430
FILING DATE: May 15, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 07/876,285
FILING DATE: April 29, 1992
APPLICATION NUMBER: U.S. 07/918,313
FILING DATE: July 21, 1992
APPLICATION NUMBER: U.S. 08/003,963
FILING DATE: January 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Richard J. Polley, Esq.
REGISTRATION NUMBER: 28,107
REFERENCE/DOCKET NUMBER: 3812-42824
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 226-7391
TELEFAX: (503) 226-9446
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4488 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double stranded
TOPOLOGY: Linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
LIBRARY: Human cDNA

POSITION IN GENOME: (of corresponding genomic gene)
CHROMOSOME/SEGMENT: 9q
MAP POSITION: 22.3
UNITS:
US-08-441-430-1

Query Match 0.7%; Score 17; DB 1; Length 4488;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2055 tgaagccaccctggaag 2071
|||||
Db 4189 TGAAGCCACCCTGGAAG 4205

RESULT 35
US-07-807-043B-8
Sequence 8, Application US/07807043B
Patent No. 5342774
GENERAL INFORMATION:
APPLICANT: Boon, Thierry, Van den Eynde, Beno t
TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/807, 043B
FILING DATE: 19911212
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5342774man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 253.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 5674 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: singular
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: MAGE-1 gene
US-07-807-043B-8

Query Match 0.7%; Score 17; DB 1; Length 5674;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1602 ggtcctgggacacctg 1618

Db 3994 GGTCTGGGACCCCTGG 4010
|||||

RESULT 36
US-08-190-411A-1
Sequence 1, Application US/08190411A
Patent No. 5541104
GENERAL INFORMATION:
APPLICANT: Chen, Yao-Tseng; Stockert, Elisabeth;
APPLICANT: Chen, Yachl; Garin-Chesa, Pilar; Rettig, Wolfgang J.;
APPLICANT: van der Bruggen, Pierre; Boon-Fallieur, Thierry;
APPLICANT: Old, Lloyd J.
TITLE OF INVENTION: MONOCLONAL ANTIBODIES WHICH BIND TO
TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR MAGE-1, RECOMBINANT MAGE-
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,411A
FILING DATE: 01-FEBRUARY-1994
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 037,230
FILING DATE: 26-MARCH-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5541104man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5354
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5674 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: MAGE-1 gene
US-08-190-411A-1

Query Match 0.7%; Score 17; DB 1; Length 5674;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1602 ggtcttgagcaccctgg 1618
 |||||||
 Db 3994 GGTCTTGAGCACCCTGG 4010

RESULT 37
 US-08-299-849B-8
 ; Sequence 8, Application US/08299849B
 ; Patent No. 5612201
 ; GENERAL INFORMATION:
 ; APPLICANT: De Placen, Etienne; Boon-Falleur, Thierry;
 ; APPLICANT: Lech, Bernard; Szikora, Jean-Pierre; De Smet, Charles;
 ; APPLICANT: Chomez, Patrick
 ; TITLE OF INVENTION: Isolated Nucleic Acid Molecules useful in
 ; TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
 ; NUMBER OF SEQUENCES: 48
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felle & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
 ; COMPUTER: IBM
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/299,849B
 ; FILING DATE: 1-SEPTEMBER-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/037,230
 ; FILING DATE: 26-MARCH-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US92/04354
 ; FILING DATE: 22-MAY-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/807,043
 ; FILING DATE: 12-DECEMBER-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/764,364
 ; FILING DATE: 23-SEPTEMBER-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/728,838
 ; APPLICATION NUMBER: 9-JULY-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/705,702
 ; FILING DATE: 23-MAY-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 5612201man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5355
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 688-9200
 ; TELEFAX: (212) 838-3884
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5674 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 ; FEATURE:
 ; NAME/KEY: MAGE-1 gene
 ; US-08-299-849B-8

Query Match 0.7%; Score 17; DB 1; Length 5674;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3994 GGTCTTGAGCACCCTGG 4010
 |||||||

RESULT 38
 US-08-560-024-1
 ; Sequence 1, Application US/08560024
 ; Patent No. 5843448
 ; GENERAL INFORMATION:
 ; APPLICANT: Chen, Yao-Tseng; Stockert, Elisabeth;
 ; APPLICANT: Chen, Yachi; Garin-Chesa, Pilar; Rettig, Wolfgang J.;
 ; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry;
 ; APPLICANT: Old, Lloyd J.
 ; TITLE OF INVENTION: MONOCLONAL ANTIBODIES WHICH BIND TO
 ; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR MAGE-1, RECOMBINANT MAGE-
 ; TITLE OF INVENTION: AND MAGE-1 DERIVED IMMUNOGENIC PEPTIDES
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felle & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
 ; COMPUTER: IBM
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/560,024
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/190,411
 ; FILING DATE: 01-FEBRUARY-1994
 ; APPLICATION NUMBER: 037,230
 ; FILING DATE: 26-MARCH-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US92/04354
 ; FILING DATE: 22-MAY-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/807,043
 ; FILING DATE: 12-DECEMBER-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/764,364
 ; FILING DATE: 23-SEPTEMBER-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/728,838
 ; APPLICATION NUMBER: 9-JULY-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/705,702
 ; FILING DATE: 23-MAY-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 5843448man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5354
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 688-9200
 ; TELEFAX: (212) 838-3884
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5674 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 ; FEATURE:
 ; NAME/KEY: MAGE-1 gene
 ; US-08-560-024-1

Query Match 0.7%; Score 17; DB 2; Length 5674;
 Best Local Similarity 100.0%; Pred. No. 87;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1602 ggtcctggcaccctgg 1618
|||||

Db 3994 ggtcctggcaccctgg 4010

RESULT 39

US-08-142-368A-8

; Sequence 8, Application US/08142368A

; Patent No. 5925729

; GENERAL INFORMATION:

; APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;

; APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;

; APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia

; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felfe & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/142,368A

; FILING DATE: 02-MAY-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/04354

; FILING DATE: 22-MAY-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/807,043

; FILING DATE: 12-DECEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/764,364

; FILING DATE: 23-SEPTEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/728,838

; FILING DATE: 9-JULY-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/705,702

; FILING DATE: 23-MAY-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 5925729man D.

; REGISTRATION NUMBER: 30,946

; REFERENCE/DOCKET NUMBER: LUD 5253.4-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5674 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

; FEATURE:

; NAME/KEY: MAGE-1 gene

US-08-142-368A-8

Query Match 0.7%; Score 17; DB 2; Length 5674;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3994 ggtcctggcaccctgg 4010

RESULT 40

US-08-967-727-8

; Sequence 8, Application US/08967727

; Patent No. 6025474

; GENERAL INFORMATION:

; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;

; APPLICANT: Van der Bruggen, Pierre; Boon-Falleur, Thierry

; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felfe & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/967,727

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/037,230

; FILING DATE: 26-MARCH-1993

; APPLICATION NUMBER: PCT/US92/04354

; FILING DATE: 22-MAY-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/807,043

; FILING DATE: 12-DECEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/764,365

; FILING DATE: 23-SEPTEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/728,838

; FILING DATE: 9-JULY-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/705,702

; FILING DATE: 23-MAY-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 6025474man D.

; REGISTRATION NUMBER: 30,946

; REFERENCE/DOCKET NUMBER: LUD 5353

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5674 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

; FEATURE:

; NAME/KEY: MAGE-1 gene

US-08-967-727-8

Query Match 0.7%; Score 17; DB 3; Length 5674;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1602 ggtcctggcaccctgg 1618
|||||

Db 3994 ggtcctggcaccctgg 4010

RESULT 41
US-08-037-230D-8
; Sequence 8, Application US/08037230D
; Patent No. 6235525
; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: van der Bruggen, Piere; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/037,230D
; FILING DATE: 26-MARCH-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6235525man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: MAGE-1 gene
; US-08-037-230D-8

Query Match 0.7%; Score 17; DB 4; Length 5674;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1602 ggtcctggagcacctgg 1618
|||||
Db 3994 ggtcctggagcacctgg 4010

RESULT 42
US-08-321-478-6
; Sequence 6, Application US/08321478
; Patent No. 5527677
; GENERAL INFORMATION:

; APPLICANT: DEGUCHI, Takeo
; APPLICANT: KINOSHITA, Moritoshi
; APPLICANT: KATSURAGI, Kiyomori
; APPLICANT: SHIN, Sadahiro
; TITLE OF INVENTION: HUMAN ARYLAMINE N-ACETYLTRANSFERASE
; TITLE OF INVENTION: GENES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/321,478
; FILING DATE: 11-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,667
; FILING DATE: 23-MAR-1993
; APPLICATION NUMBER: JP 64669/1992
; FILING DATE: 23-MAR-1992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6464 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 723..1595
; FEATURE:
; NAME/KEY: exon
; LOCATION: 717..1936
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1794..1799
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1800..1805
; US-08-321-478-6

Query Match 0.7%; Score 17; DB 1; Length 6464;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 982 aggtaccgaaggaagc 998
|||||
Db 5916 aggtaccgaaggaagc 5932

RESULT 43
US-07-853-913-1
; Sequence 1, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator Of
; TITLE OF INVENTION: Neuroepithelial Tumors
; NUMBER OF SEQUENCES: 4

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AAAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11236 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-853-913-1

Query Match          0.7%; Score 17; DB 1; Length 11236;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1827 cctcagaagaaggctg 1843
Db 3466 CTTGAGGAGGAGGCTG 3482

RESULT 44
US-08-781-891-207
Sequence 207, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-Eu
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 207:
SEQUENCE CHARACTERISTICS:
LENGTH: 29604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-207

Query Match          0.7%; Score 17; DB 3; Length 29604;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 475 aaagaaatagactgcg 491
Db 14826 AAAGGAATGAACTGCG 14842

RESULT 45
US-09-335-409-1/c
Sequence 1, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335,409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match          0.7%; Score 17; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1407 cggccagcccccag 1423
Db 20332 CGGCCAGCCCCAGCAG 20316
```

Search completed: October 28, 2001, 21:03:46
Job time: 5450 sec

• • •

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

[illegible]

Qy	241	aacggatctcttctaactcytggaagaagcgcttcaggaatccatgacgaagacacaagta	300
Db	262	AACCGGTATCTCTTAACTGTGGAGAAAGCGCTTCAGAGACTCATGCAAGACACAAAGTTA	321
Qy	301	aaggttcgtcgcttgagacaacatattccctgcacgaatgacgtgctcaatgtttggygc	360
Db	322	AAGGTTCGTCCGCCCTGGAACAACATATTTCTGCACAGAAATGCACGTGTAAAGTTGGGGCC	381
Qy	361	ttaatgtgaatgattcttactttaaagaaacccggcttccaaagtgtgtacttctggga	420
Db	382	TTAAATGTGAATGATTTCTTACTTTAAAGGAAACCGGGCTCCAAAGTGTGTACTTTCGGA	441
Qy	421	ccctcacacatggaanaaaatacctcgaaagcaatcaaatattcttggttcatgtgaagga	480
Db	442	CCTCCACAACTGGAGAAATTAACCTGCAGAACATTCAAAATATTTCTGTGCTCAATTGAAGGA	501
Qy	481	atagaacatgycgtctgtgycggcccaactctgcgcccaagaatacgaagatgaacacatgacgtt	540
Db	502	ATAGAAGATCGGCTGTGTGGCGCCCACTCTGCCACGAATACGAGGATGAACACATGACAGTT	561
Qy	541	taccggtatcccatatcacacatgtaacaagagaggggaaagcccaacatgacagatcca	600
Db	562	TACCGATTCCTCCATACAAAGTAACAGTAACAGGAGGGAAGCAACCATGGCAGAGTCCA	621
Qy	601	gaaagcgctctcaagcaggtcagtcagtcacaagatcttca	639
Db	622	GAAAGCCTCTCAGACGGCTCAGTCTCAGACGAGATCTTCA	660
RESULT	3		
LOCUS	AU141334	666 bp	mRNA
DEFINITION	AU141334	THYR01 Homo sapiens cDNA clone THYR01000421 5', mRNA	25-OCT-2000
ACCESSION	AU141334	Sequence.	
VERSION	AU141334		
KEYWORDS	AU141334.1	GI:11002855	
SOURCE		EST.	
ORGANISM		human.	
		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE		Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,	
AUTHORS		Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and	
		Isogai,T.	
TITLE		HRT human cDNA project	
JOURNAL		Unpublished (2000)	
COMMENT		Contact: Takao Isogai	
		Genomics Laboratory	
		Helix Research Institute	
		1532-3 Yana, Kisarazu, Chiba 292-0812, Japan	
		Tel: 81-438-52-3951	
		Fax: 81-438-52-3952	
		Email: genomics@hri.co.jp	
		HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix	
		Research Institute; cDNA library construction; Department of	
		Virology, Institute of Medical Science, University of Tokyo, and	
		Helix Research Institute	
FEATURES		Location/Qualifiers	
Source		1. .666	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="THYR01000421"	
		/clone_1db="THYR01"	
		/tissue_type="thyroid gland"	
		/note="Vector: pME18SFl3"	
BASE COUNT	158 a	186 c	190 g
ORIGIN			3 others
Query Match	24.5%	Score 609;	DB 108; Length 666;
Best Local Similarity	100.0%;	Pred. No. 0;	


```

ACCESSION   BF969043
VERSION     BF969043.1
KEYWORDS    GI:12336258
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 944)
AUTHORS    NIH-MGC http://cncg.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LHAM995 row: d column: 18
            High quality sequence stop: 591.
            Location/Qualifiers
                source          1..944
                                /organism="Homo sapiens"
                                /db_xref="taxon:9606"
                                /clone="IMAGE:4357985"
                                /clone_lib="NIH_MGC_84"
                                /tissue_type="adrenal cortex carcinoma, cell line"
                                /lab_host="DH10B (phage-resistant)"
                                /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site:1:
                                NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT
                                primed. Average insert size 1.229 kb. Library enriched for
                                full-length clones and constructed by Life Technologies.
                                Note: this is a NIH_MGC Library."
BASE COUNT   200 a      266 c      264 g      214 t
ORIGIN
Query Match      24.1%; Score 597; DB 172; Length 944;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 gttggagacagctgcgtcccatcatgtctgctgtgtcaagagcggaatacatcact 855
    |||||||
DB 1 GTTGGAGACACTGCCATCGCTCCCATCATGTGCTGTCAAGAGCGGAATACATCACT 60
    |||||||
QY 856 catgaaggaagaagatttggctgaagagctgttactctccagatcctgtgtctgt 915
    |||||||
DB 61 CATGAAGGAAGAAGATTGGCTGAAGAGACTGTGTACTCCAGATCCTGTGCTGCT 120
    |||||||
QY 916 ttgtgtgtgtagaatgtccagatgaagatcatcattcaaccatctgtgaatgccac 975
    |||||||
DB 121 TTGTGTGTGTGAATGTCCAGATGAAGCTTATTCACACCCTGTGTGGAATGCCACC 180
    |||||||
QY 976 ttccagaggtaccagaagaagacagatgcctccgtgtgctgtgtgttccatgtgccca 1035
    |||||||
DB 181 TTTCAGAGGTACCAAGGAAGAGAGATGCCCGCTGTGGTGTTCATATGGCCCCCA 240
    |||||||
QY 1036 gcatcgtgtctgtgtgacagcaggtaccagcagtgatgataagaggttttggtgctgacac 1095
    |||||||
DB 241 GCATCTGTGCTTGTGACACAGCTACCAACATGTGATGAGAGGTTTGAGCTGACACC 300
    |||||||
QY 1096 cagcacttgctcgtgaatggaactgtgctcagttcaacaacctgcgaagcaagaatt 1155
    |||||||
DB 301 CAGCACTTGTGCTCTGAATGAGAACTGTGCTCAGTTCAACAACCTTGCACCAAGATT 360
    |||||||
QY 1156 caaacccagctcaactcatcaccgagacatctccctctgtcaccagtttcgctgt 1215
    |||||||
DB 361 CAAACCCAGCTCAACTCATCCACCGGAGCATCTCCCTGTGTCAACATTTCCGCTGT 420
    |||||||
QY 1216 aagaagaaggagcccaacctcagttgtgcccatgtgttcagggtgaatgctctccaagtac 1275
    |||||||

```

```

DB 421 AAGAAGAGGGCCCCACACCTCACTAGTGTGCCCATGCTTCAGGGTGAATGCCCTCAAGTAC 480
QY 1276 cagctccgtcccaaggaggtgacagaagatgccattactgtgcaactctgaagaa 1335
    |||||||
DB 481 CAGCTCCGTCCAGAGAGAGAGTGGCAGAGGAGATGCCATTATTACTTGCATCTGAGAGA 540
    |||||||
QY 1336 tcatagtgaagcgtgtcagcttcccaacttccagacagagcgtgtcagaagatcag 1392
    |||||||
DB 541 TTGATAGTTGAGCGGCTGCACTTCCCACTTCAGCAGAGAGCTGCAGAGTACAGG 597
    |||||||

RESULT 6
BE383336      664 bp      mRNA      EST      21-JUL-2000
LOCUS        601298249F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5',
DEFINITION   mRNA sequence.
ACCESSION   BE383336
VERSION     BE383336.1 GI:9328701
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
            1 (bases 1 to 664)
            NIH-MGC http://cncg.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
            Plate: LHAM313 row: a column: 13
            High quality sequence stop: 662.
            Location/Qualifiers
                source          1..664
                                /organism="Homo sapiens"
                                /db_xref="taxon:9606"
                                /clone="IMAGE:3628308"
                                /clone_lib="NIH_MGC_19"
                                /tissue_type="neuroblastoma"
                                /lab_host="DH10B (phage-resistant)"
                                /note="Organ: brain; Vector: pORF7; Site:1: XhoI; Site:2:
                                EcoRI; cDNA made by oligo-dT priming. Directionally
                                cloned into EcoRI/XhoI sites using the following 5'
                                adaptor: GGCACGAG(G). Library constructed by Ling Hong
                                in the laboratory of Gerald M. Rubin (University of
                                California, Berkeley) using ZAP-cDNA synthesis kit
                                (Stratagene) and Superscript II RT (Life Technologies).
                                Note: this is a NIH_MGC Library."
BASE COUNT   154 a      199 c      174 g      137 t
ORIGIN
Query Match      23.8%; Score 590; DB 166; Length 664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1093 acccagcacttgctcgtgaatggaactgtgctcagttcaacaacctgcgaagcaaga 1152
    |||||||
DB 1 ACCCAGCACTTGTGCTCTGAATGAGAACTGTGCTCAGTTCAACAACCTTGCACACCAAG 60
    |||||||
QY 1153 atcaaacccagctcaactcatcaccgagacatctccctctgtcaccagtttcgac 1212
    |||||||
DB 61 ATTCAAAACCAGCTCAACTCATCCACCGGAGCATCTTCCCTGTGTCAACATTTCCCGC 120
    |||||||
QY 1213 ttrtaagaagagagcccaacctcagttgtgcccatgtgttcagggtgaatgctctccaag 1272
    |||||||
DB 121 TGTAAAGAGAGGGCCCCACACCTCACTAGTGTGCCCATGTTGAGGATGAATGCTTCTCAAG 180
    |||||||

```


High quality sequence stop: 736.
location/Qualifiers

FEATURES
source 1..920

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583679"
/clone_lib="NIH_MGC-15"
/tissue_type="adenocarcinoma cell line"
/note="Organ: colon; Vector: pOPF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(9). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 196 a 261 c 272 g 190 t 1 others
ORIGIN

Query Match 22.6%; Score 561; DB 153; Length 920;

Best Local Similarity 99.9%; Pred. NO. 4.6e-285;

Matches 681; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 916 ttgtgtggtgtagaattgcatgaaagcttcataccacatctgtgagaatgccac 975
|||||
Db 21 TTTGTGTGTGATGATTCAGATGAAAGCTTATTCACCCATCTGTGAATGCCACC 80
QY 976 ttccagaggtaccagaagaagcagatgccccgtgacctgtgtgttcacatgccccca 1035
|||||
Db 81 TTTCAGAGGTACCAAGGAAGGCGAGATGCCCGCTGGCTTGTGATTCACATGGCCCCCA 140
QY 1036 gactctgtgcttctgagcagcaggtaccagatgagatgagaggtttggccctgacacc 1095
|||||
Db 141 GCATCTGTGCTTGTGTGACACAGGTACAGAGTGGAGAGGTGGCGCTGCACACC 200
QY 1096 cagcacttggctcctgaaatgaaactgtgctcagttcacacaacctctgcagccacaagaatt 1155
|||||
Db 201 CAGCACTTGTGCTCGATGAGAACTGTGCTCAGTTTCAACAACCTGCGACCAACAATTT 260
QY 1156 caaacccagctcaactcaccacccggagacattcccccctgtctaccagttccgctgt 1215
|||||
Db 261 CAACCCAGGTCAACCTCAATCCACCGGACATCTTCCCTGCTCACACAGTTTCCGCTGT 320
QY 1216 aagaagaagggcccccacccctcagtgctgcccagtgtaaggtgaatgccctcccaagttac 1275
|||||
Db 321 AAGAGAGAGGGCCCCACCTCACTGTGCCCCATGTGAGGTGATGCCCTCTCAAGTAC 380
QY 1276 cagctccgtcccaaggaaggtgagcaaggaatgcatltaacttgcacatcctgagaa 1335
|||||
Db 381 CAGCTCCGTCCTCCAGAGAGGAGTGGAGAGGATTCATTTACTTCCAAATCCGAGAGAA 440
QY 1336 ttcatagttagagcgctgacgcttcccaacttccagcagagcggtgacagaggaagg 1395
|||||
Db 441 TTTCATGATTTAGGCGCTGCGCTTCCCACTTCCAGCAGAGCGTGCAGGAGTACAGAGAG 500
QY 1396 agtggcagagaagggcccaagccagcagagaagaagtaagtaaccagaataatcatttc 1455
|||||
Db 501 AGTGGCCAGAGACGGCCCAAGCCCAAGCAGAGAAAGAGTACCCAGAAATCATCTTC 560
QY 1456 cttygaacagaggtctgcatcccgatgaagatlcgaatgctcagtgccacactgtgtaac 1515
|||||
Db 561 CTTGGAAGAGAGGTCTG-CATTCGCGATGAAGATTGCAATGTCAGTGCACACTTGTGAAC 619
QY 1516 ataaagcccgacacgctctcgtactactgagactgtgtgtgagggacacatttggcagctgtgc 1575
|||||
Db 620 ATTAAGCCCCACACAGTCTGTCTACTGAGACTGTGTGAGGGCACATTTGGGCAAGCTGTGC 679
QY 1576 cgtcattacaggaagcaggtg 1597
|||||
Db 680 CGTCAATTACGAGACCAAGGTGG 701

RESULT 9

LOCUS BF434169

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BF434169 553 bp mRNA EST 29-NOV-2000
7094704.x1 NC1-CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644670 3'
similar to TR:09V5J4 Q9V5J4 CG3298 PROTEIN.; mRNA sequence.
BF434169
BF434169.1 GI:11446441
EST.
human.
Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 553)
NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NC1-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
infoimage.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 507.
Location/Qualifiers
1..553

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3644670"
/clone_lib="NC1-CGAP_Ov18"
/tissue_type="fibrothecoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pRT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dt) primer [5'
TGTTCACATCTGAGAGTGGAGCGGCGGCGGCAATTTTGTGTGTGT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 143 a 150 c 153 g 107 t

ORIGIN

Query Match 22.3%; Score 553; DB 148; Length 553;

Best Local Similarity 100.0%; Pred. NO. 7.7e-281;

Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 cggcgcgcaaggaaccgctgctgcaacctgcaagcagagaagagcgagaccgtcgagg 150
|||||
Db 1 CGGCGCGGCAAGAGACCGCTGCGGCACTGCGACGAGAGAGAGCGGACCGTGGGG 60
QY 151 tgcctcgcgccgccaacacacgctgtacactgcaagtgtgtgcaagcggtgtagcggagactcg 210
|||||
Db 61 TGCCTCGCGCGCCCAACACCGTGTACTGTCAGTGTGTGCGACGCGGTACCGCGACTCG 120
QY 211 gggcgcgcgctcagtgctctcgcggttaaacgcggtatctcttaactgtgagaagc 270
|||||
Db 121 GGGCGCGGCTTACGCTTCTCCGAGTTCAACCGGATCTTCAACTGTGAGAGAGG 180
QY 271 gtccagagactcagcaggaagcaagttcaagttcgtcgtcgtggaacaacatctcctg 330
|||||
Db 181 GTTCAGAGACTCATGACGAGGACACAGTTAAAGTTGCTGCTCGTGCACAAATATTCCTG 240
QY 331 acaagaaatgactgtctaaatgttgggggttaagtgaatgattcttacttaagaa 390
|||||

Db 241 ACACGAATGCATGCTTAATGTTGGGGCTTAAGTGAATGATCTTACTTTAAAGCA 300
OY 391 accgggtccaaagtgtgtactttcttgacctccacaactggaataactctgaagca 450
Db 301 ACCGGCTTCCAAAGTGTACTTCTGACCTCCACAACGGAATAATACCTCAACACA 360
OY 451 atcaaatattcttcgtccattgaaagaataagacgtgctgctcgcccaactctgc 510
Db 361 ATCAAAATATTTCTGTGTCATGAAAGAAATAGAACTGGCTGTCCGCCCTCTGCC 420
OY 511 ccagaatacgsagatgaacaacatgacagttaccagaatcccaatacacagtgaagag 570
Db 421 CCAGATATGAGATGAAACCATGACAGTTACCAAGATCCCATACACAGTGAACAGAG 480
OY 571 aggggaagcccaacacatgacagttccagaagagccttcacagagctcaagtcagag 630
Db 481 AGGGAAAGCACCAACCTGGCAGAGTCCAGAAAGCCTTCAGCAGGCTCAGTCCAGAG 540
OY 631 cgatctcagact 643
Db 541 CGATCTCAGACT 553

RESULT 10
BG396395 827 bp mRNA EST 12-MAR-2001
LOCUS 60245932F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4581700 5',
DEFINITION mRNA sequence.
ACCESSION BG396395
VERSION BG396395.1 GI:13289941
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 827)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LML at: image.llnl.gov
Plate: LCM1302 row: n column: 05
High quality sequence stop: 779.
Location/Qualifiers

FEATURES
Source
1. 827
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4581700"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site: 1: XhoI; Site: 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 205 a 217 c 226 g 179 t
ORIGIN

Query Match 21.9%; Score 543; DB 153; Length 827;
Best Local Similarity 100.0%; Pred. No. 1.5e-275;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 707 actctccctgctcgaacttcatctgtaagcttcaacttaagaaggaactcttgg 766
Db 22 ACTCTTCCCTGGTGTGCTTACTTTCACTTAAAGAGAGAACTTCTTGG 81
OY 767 tgcctaaagcaagagatggtgctcccaagttggacagctgcacatgctcccatatg 826
Db 82 TGCTCAAGCAAGAGATGGGCTCCAGTGTGGACAGTGCATCGCTCCATCATTTG 141
OY 827 ctgctgtaagaagaggaagaagcatcaactcaatgaagaagagatlltggctgaagagc 886
Db 142 CTGCTGTAAAGACGGAAGAACATCACTCATGAAGAGAGATTTGGCTGAAGAGC 201
OY 887 tgtgtaacctcccaagatcctgtgctgttctgtgtgtgtgtaagatgacagatgaagct 946
Db 202 TGTGTACTCTTCAGATCTGTGTGCTGCTTTTGTGTGTGTGAATGTCCAGTGAAGCT 261
OY 947 tcaatcaaccatctgtgagaatgcaacttccagagtlacaaagaaagcagatggcc 1006
Db 262 TCATCAACCCATCTGTGAATGACACCTTTACAGAGTACCAGAAAGGAGATGCC 321
OY 1007 ccgtgacctgtgtgtacatgagcccaagcatctgtctgtgtgacagcagtaaccagc 1066
Db 322 CCGTGGCCTTGTGTGTCATATGGCCCAAGCATCTGTCTGTGTGACAGAGTACAGC 381
OY 1067 agtgaatgagaggtttgtgacctgacaccagcaacttgctctgaatgaagactgtgct 1126
Db 382 AGTGATGAGAGGTTGTGGCTGTACACCAAGCATTTGTCTGATGATGAGATCTGTGCT 441
OY 1127 cagttcaaaccttcgcagcacaagaattcaaacccaagctcaactcatccaccggaca 1186
Db 442 CAGTTCACACACTTTCGACGCCACAAAGATTCAAAACCCACTCAACTCATCCACCGGACA 501
OY 1187 tcttcacctgtcacaacagtttcgctgtgaagaagagggcccaacctcagttgtccca 1246
Db 502 TCTTCCCTGCTGTCCAGTTTCCGTGAAGAGAGGCGCCACCTCAGTGTGCCCA 561
OY 1247 tgg 1249
Db 562 TGG 564

RESULT 11
BG335963 614 bp mRNA EST 27-FEB-2001
LOCUS 602404708F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4542553 5',
DEFINITION mRNA sequence.
ACCESSION BG335963
VERSION BG335963.1 GI:13142401
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 614)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LML at:
<http://image.llnl.gov>
Plate: LCM1222 row: o column: 02
High quality sequence stop: 614.
Location/Qualifiers

FEATURES
Source
1. 614
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4542553"

QY	2079	ggaagagggaagcagcgtgggaagaaagcacaacagcagcagcgtcccaagccatcagcgtggagat	2138
Db	600	GGAAAGAGGAGCAGTCGTGGAAAGACACACAGCACAACGTCCTCCAGCCATTCAGCTGGGAGT	659
QY	2139	gcgagatgaacgcgcgaggttcattatgctgtcaacacttcagc	2178
Db	660	GGCGATGAACGGGAGGTTCATATATCTGTAACCACTTCAGC	699
RESULT	13		
LOCUS	BF196304		
DEFINITION	BF196304	526 bp mRNA	EST
ACCESSION	BF196304	7n68a10.x1 NCI CGAP OV18 Homo sapiens cDNA clone IMAGE:3569586	03-NOV-2000
VERSION	BF196304.1	similar to TR:Q9V5J4 Q9V5J4 CG3298 PROTEIN. ;, mRNA sequence.	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 526)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapsb-remail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prentge, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center		
FEATURES	Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/JLNL, send email to: info@image.lnl.gov		
SOURCE	High quality sequence stop: 491.		
	Location/Qualifiers		
	1..526		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3569586"		
	/clone_1lb="NCI CGAP OV18"		
	/tissue_type="fibrotheloma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: ovary; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGGCGGCGCGGACATATTTTATTTT 3']; Site_2: Eco RI sites of the modified p773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	136 a 142 c 147 g 101 t		
ORIGIN			
Query Match	21.2%; Score 526; DB 145; Length 526;		
Best Local Similarity	100.0%; Prid. No. 1.5e-266;		
Matches	526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	90	gcggcgcgcgcgaagcccgctgcgccacctgctgcgcggaaggaagcgcgcgcgtcggg	149
Db	1	GGGGCCGGGCAAGGACCCGCTGGCGACCTGGCAACGGAGAAAGCCGGACCTCGGG	60
QY	150	gtgctcgcgcgcgcgaacacacgctgtacctgcgaagtgtgtgcagcgcgtgtacgcgcgactc	209
Db	61	GTGCTCGGGCGGCCCAAAACACGCTGTACCTGACAGGTGTGTGCGCAGCGGAGCCGGGACATC	120
QY	210	gggcgcgcgcgtctcagctctcttcgcgagttcaaccggtatctcttcaactgtgtgaagaag	269

```

Db      121 GGGGCGCCGCCTTACAGTCCTTCGCCAGTTCAACGCCGATCATTCTTAACA CTGTGGAGAAG 180
QY      270 cgttcagagactcatgcaggacacaagttaaaagtgctcgctcgtggaacataatcc 329
       |||||
Db      181 CGTTCAGAGACTCATWGCAGGAGCACAAAGTTAAGGTTTGCTGCCTGGACACACATATTTCT 240
QY      330 gaccgcaatlgacctgyrctaattgttggsgggttaagtgaatgatctttactttaaga 389
       |||||
Db      241 GACACGAATGCACGTCCTAATGTGGGGGCTCTTAA GTTAGAATGATTTCTTTAAAAGA 300
QY      390 aaccgggcttcctcaaagltgctactcttcygaacctccaactcygaaaatbcctcgaag 449
       |||||
Db      301 AACCGGGCTTCCAAGTAGTGTTACTTTTCGACCTTCACAACTGGAAAAATACCTGGAAGC 360
QY      450 aaaccaaatatcttccttgtlccatlgaaagaatagaaactgycgtgtycgccccactctgc 509
       |||||
Db      361 AATCAAATAATTTCTTGCTGCATTTGAAGAAGATAGAACATGGCTGTGGCGCCCACTGTGC 420
QY      510 cccagaatacgaagatgaaacccaatgacagtttaccaagatcccacatacacagtgaacag 569
Db      421 CCCAGAAATACAGAGATGAMAACCATGACAGCTTTAC CAGATCCCACATACACAGTACAAG 480
QY      570 gaggggaaagcaccacccatgcsagaatgccagaagcgctctaac 615
Db      481 GAGGGGAAAGCACCACCATGGCAGATGCCAGAAAGCGCTCTCAGC 526

RESULT_14
LOCUS     AU138595             696 bp    mRNA           EST            25-OCT-2000
DEFINITION AU138595 PLACE1 Homo sapiens cdna clone PLACE1008903 5', mRNA
sequence.
VERSION   AU138595
KEYWORDS  AU138595.1 GI:11000116
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS   I. (bases 1 to 696)
Orita,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Iisogai,T.
TITLE     HRI human CDNA project
JOURNAL   Unpublished (2000)
COMMENT   Contact: Takao Iisogai
          Genomics Laboratory
          Helix Research Institute
          1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
          Tel.: 81-438-52-3951
          Fax: 81-438-52-3952
          Email: genomics@hri.co.jp
          HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
          Research Institute; cDNA library construction; Department of
          Virology, Institute of Medical Science, University of Tokyo, and
          Helix Research Institute.
FEATURES             Location/Qualifiers
     source           1..696
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1008903"
                     /clone_id="PLACE1"
                     /tissue_type="placenta"
                     /note="Vector: pMEBSPFL3"
BASE COUNT  163 a        192 c        208 g        130 t        3 others
ORIGIN
Query Match              21.2%; Score 525; DB 108; Length 696;
Best Local Similarity 100.0%; Pred. No. 5,1e-266;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Db	73	ATGTGGGCGCTTGGCTCGTGCTGGGTTCCGGCGGCCAGCACCAATGTCGACGGAGGC	132
Oy	61	accataatcgcaaggcacccgcgccgcgacggcgcgcaagaagaccgcgtctgcgacctg	120
Db	133	ACCATAATCGCAGCAGCACCCGCCGCCGACGCGGCCGCAAGAGAACCCGTCGCGACCTG	192
Oy	121	cgcacgcgagaaagacgcgggaacgtcgggtggtcttcggcgggcccacaacacgttactgt	180
Db	193	CGCACGCGAGAAACAACGGGGACCGTCCGGGGGTCTCCGGCGGCCCAAACACCGTGTACTCG	252
Oy	181	caggtgtgttgcaagcggtctagccccgggaaccgcggcgccgcgcctctacgtttccgaattc	240
Db	253	CAGGTGTGTGCGCAGCGGGTAGCCGGGACTCCGGGCCCCGCGCTCATGCTTCTCCGAGTTC	312
Oy	241	aaccggtatctcttaactgttggaagaagcgcttcagaaactcatgatcagagacacaagtta	300
Db	313	AACCGGTATCTCTTAACGTGTGGAGAAGCGTTGAGAACATCATGACAGAGACACACAAGTTA	372
Oy	301	aaggttgtctgcgcctgcgacaacatatctctgcagacgaalgtacgtgtctaattgttggggc	360
Db	373	AAGGTGTGCTCGCGCTCGAGAACATATTCTTGACACGAATGCACTGGTCTAATGTGGGGGC	432
Oy	361	ttaagtgtaatgattcttactttaagaagaaccgggcttccaagtgtgtacttttggga	420
Db	433	TTAAGTGTGATATTCTTACTTTAAAGGAACCCGGGCTCCAAAGTGTTACTTTCTGGA	492
Oy	421	cctccacaactgtgaanaaacctctgaagacaatcaaatatttctgtgtccattgaagga	480
Db	493	CCTCCACAACTGTGAAAAATTAACCTCGAACAAATCAAAAATATTTTCTGTGTCATTGAAGGA	552
Oy	481	atagaactgtgtctgtcgggccccaactctgcaccaagaatacagaagat	525
Db	553	ATAGAAGCTGGCTGTGGGCCCCACACTCTGCCACAGATATCGAGGAT	597
RESULT	15		
LOCUS	BE794311		
DEFINITION	6015591442.1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:394538 5', mRNA sequence.		
ACCESSION	BE794311		
VERSION	BE794311.1 GI:10215496		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 688)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: c9apbs-remail.nih.gov Tissue Procurement: DCPD/DTP cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: Image.LNL.gov Plate: LNC804 row: c column: 11 High quality sequence stop: 688.		
FEATURES			
source	1..688	Location/Qualifiers	

[illegible]

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Louis Staudt, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM952 row: c column: 07
High quality sequence stop: 705.
Location/Qualifiers
1. .960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4340286"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: PCMV-Spore; site_1: NotI; site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 230 a 286 c 276 g 168 t
ORIGIN

Query Match 20.9% Score 519; DB 169; Length 960;
Best Local Similarity 99.6% Pred. No. 7.7e-263;
Matches 669; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1362 caattccagacagacgtgacagagatgacagagagtcgacagacgcccacccagc 1421
|||||
DB 1 CAACATCCAGACAGACCGTCGAGAGTACAGAGAGTCCGACGAGCGCCACCGC 60
|||||

QY 1422 agagaaagacagacagacagacagacagacagacagacagacagacagac 1481
|||||
DB 61 AGAGAAAGAGTACAGTACCCAGAAATCATCTCTTGAGACAGGCTCGCATCCGAT 120
|||||

QY 1482 gaagatcgaagatgcaatgacagacagacagacagacagacagacagacagac 1541
|||||
DB 121 GAAGATTCGAAATGTCATGTCACACTTGTTCACATTAAGCCCCGACAGTCTGCTACT 180
|||||

QY 1542 ggaactgtgtagagacacatttgacagacgtgacacatttgacagacagacag 1601
|||||
DB 181 GGACTGTGTAGGACAGTGTGGGACACTGTGCGCTATTAGGAGACAGAGTGGACAG 240
|||||

QY 1602 ggtcctgggcaacctgtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1661
|||||
DB 241 GGTCCTGGGCAACCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
|||||

QY 1662 ctggcgaagatctgt 1721
|||||
DB 301 CTGGCAAGATCTGT 360
|||||

QY 1722 cccctgt 1781
|||||
DB 361 CCCCTGT 420
|||||

QY 1782 gtgcgaagagtcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1841
|||||
DB 421 GTGCAGAGAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
|||||

QY 1842 tgaagatccagtcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1901
|||||
DB 481 TGAGATCTCCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
|||||

QY 1902 ggaagagttcagacacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1961
|||||
DB 541 GGAAGAGTTTCAAGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
|||||

QY 1962 gcaacccctgt 2021
|||||
DB 601 GCACACCTGT 660
|||||

QY 2022 ccgatgt 2083
|||||
DB 661 CCGGATGGGGAA 672
|||||

RESULT 17
AU143668
LOCUS AU143668 670 bp mRNA EST 25-OCT-2000
DEFINITION AU143668 Y79AA1 Homo sapiens cDNA clone Y79AA1002297 5', mRNA
sequence.
ACCESSION AU143668
VERSION AU143668
KEYWORDS AU143668.1 GI:11005189
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 670)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.
HRI human cDNA project
Unpublished (2000)
CONTACT: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. .670
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Y79AA1002297"
/clone_lib="Y79AA1"
/cell_type="retinoblastoma"
/cell_line="Y79"
/note="Vector: pME18SFL3"
BASE COUNT 157 a 190 c 193 g 127 t 3 others
ORIGIN

Query Match 20.4% Score 507; DB 108; Length 670;
Best Local Similarity 99.7% Pred. No. 1.7e-256;
Matches 607; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtggcgcttctgt 60
|||||
DB 31 ATGTGGGCGCTTGT 90
|||||

QY 61 accatatcgagagaccccgccgagagagagagagagagagagagagagagagagag 120
|||||
DB 91 ACCATATGCAAGGACCCGCGCGCGGACGCGCGCGGACGAGACCCCTGGGCACTG 150
|||||

QY 121 cgcacgacagaaacgc 180
|||||
DB 151 CGCACGACAGAAACGCGGAGACGCGGAGGCTCGGCGGCGGCGGCGGCGGCGG 210
|||||

QY 181 caggt 240
|||||
DB 211 CAGGT 270
|||||

QY 241 aacggtatcttcaactgt 300
|||||

```
|||||
Db 271 AACCGATCTCTTCACTGTGAGAGGCGTTTCAGACTCATGCGAGCAACAATTGA 330
QY 301 aaggtgctgcctgagcaacatactcctacagcaatgacggtcttaattgtg99gc 360
Db 331 AAGGTCTCTGCTGGACGACATATTCCTCAGACGAATGCACGTCTAATGTTGGGGC 390
QY 361 ttaagtgaatgattcttactttaagaacccggtctccaaagtgtactcttga 420
Db 391 TTAAGTGAATGATTCTTACTTTAAAGGAACCGGCTTCCAAAGTGTAATTTCTGGA 450
QY 421 cctccacaactggaataatcctccgaagcacaataatcttctgtcattgaaagga 480
Db 451 CCTCCACAATGTGAAAAATATCTCGAAGCAATCAAAATATTTCTGCGCCATTGAAAGGA 510
QY 481 atgaactgctgtgcgcccactctgcgccagaatacaggaatgaagaactgaagtt 540
Db 511 ATGAGACTGGCCGTGGCGCCCACTGTGCCCCAGAAATACAGATGAACCATGACAGTT 570
QY 541 taccagatcccaatacacagtgacagagaggggaaagcacaacatgtgcagatcca 600
Db 571 TACCAAGATCCCATACACTGAAACAGAGAGGGAAGAACCAACCATGCGAGTCCA 630
QY 601 gaaagcct 609
Db 631 GAAAGGCT 639
```

```
RESULT 18
BF057321 506 bp mRNA EST 16-OCT-2000
LOCUS 7K19c01.x1 NCI-CGAP.Ov18 Homo sapiens cDNA clone IMAGE:3475944 3'
DEFINITION similar to TR:Q9V5J4 Q9V5J4 CQ3298 PROTEIN.;, mRNA sequence.
ACCESSION BF057321
VERSION BF057321.1 GI:10811217
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento
Soares, Ph.D. CDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 477.
Location/Qualifiers
1. 506
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP.Ov18"
/lisse_type="fibrotheoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pTR73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dT) primer [5'
TGTTCACATCTGAGGTGGAGCGCGCGCGCATTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pTR73 vector. Library
went through one round of normalization, and was
```

```
BASE COUNT 133 a 133 c 139 g 101 t
ORIGIN
Query Match 20.4%; Score 506; DB 143; Length 506;
Best Local Similarity 100.0%; Pred. No. 5,6e-256;
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 gctgcgacacctgcgacgcgagagaagcggaccctcgagggtgtctcgcg99ccaaa 167
Db 1 GCTGCGGACCTGCGCACGCGAGAGAAAGCCGACCGTCCGGGTGCTCCGGCGCCCAA 60
QY 168 caccgltaccttgcaagttggtgagcagcgggtaacggagactcggcgccgctcaagt 227
Db 61 CACCGGTACTCTGACAGGTGTGGCAGGGGTAGCCGGACTCGGGCGCCGCTTACGT 120
QY 228 ctctccgagttcaaccggtatctcttcaactgtgagaagcggttcagaagactatga 287
Db 121 CTCTCGAGTTCAACCGGTATCTCTTCAACTGTGGAAGGGGTTCCAGAGCTCATGCA 180
QY 288 gggacacaagttaaaggtgtcgtccgtgcgacacatactctgcagcaatgactgtgc 347
Db 181 GGAGCAACAGTTAAAGSTTGTCTGCTGACACATATTTCTTACACAGATGCACTGTG 240
QY 348 taatgttgaggcttaagltgaatgattcttactttaaagaacccggctccaaagt 407
Db 241 TAATGTTGGGGGCTTAAGTGAATGATTTCTTAAAGAAACCGGCTTCCAAAGTG 300
QY 408 tgtacttctgacacctccacaactggaataatcctcgcgaacataataattctg 467
Db 301 TGTACTTTCGTGACCTCCACACTGGAATAATCTCGAAGCAATCAAAATATTTCTTG 360
QY 468 tccattgaaaggaatgaactgtctgctgcgcccactctgcgccagaatacagagatga 527
Db 361 TCCATTGAAGGAATACACTGTGCTGTGCGGCCCACTGTGCCCAATACAGATGA 420
QY 528 aacctgacagtttaccagatccccaatacacagtgacagagaggggaaagcacaacc 587
Db 421 AACCATGACAGTTTACACAGATCCCATACACTGAAACAGAGAGGGAAGCAACCAAC 480
QY 588 atggcagagttccagaagaagcctctca 613
Db 481 ATGCACAGTCCAGAAAGGCTCTCA 506
```

```
RESULT 19
BF794577 953 bp mRNA EST 12-JAN-2001
LOCUS 602255843F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4339187 5',
DEFINITION mRNA sequence.
ACCESSION BF794577
VERSION BF794577.1 GI:12099631
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://mgc.ncbi.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Contact: cgabs-r@mail.nih.gov
Tissue Procurement: Louis Staudt, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM9949 row: e column: 12
High quality sequence stop: 704.
```


source 1. 823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3876223"
/clone_1lb="NH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6, site_1: NotI;
site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 185 a 219 c 259 g 160 t
ORIGIN

Query Match 18.7%; Score 465; DB 138; Length 823;
Best Local Similarity 100.0%; Pred. No. 2.8e-234;
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1628 ttgttcccaactgcaagcagatcacacacagggcttgccagatcttctgtagagag 1687
|||||
DB 75 TTGTGTCACACCTGACGACAGATCACACAGGGCTTGCCAGATATCTTGCTGCAGAGAG 134
|||||
QY 1688 aacggccttggaactcttgggaagccgcttcaaccttggctggctggcccaacc 1747
|||||
DB 135 AACCGGCTTGGAATCTTTGGGAAGCGGCTTACCCCTTGCTGCTGGTGGCCCAACC 194
|||||
QY 1748 agctcaagcctgctcagcagatcaccaaacagctccagagtgctctgacacatca 1807
|||||
DB 195 AGCTCAAGCCTGGCTCAGACAGTACCAACACAGTCCAGAGAGTCTGCACACATCA 254
|||||
QY 1808 gtagtattcctgccaatgctcctcaagaagggctgtagatctcagctcagctgga 1867
|||||
DB 255 GATGATTCCTGCGCAATGCTTACAGAAAGGGGCTGAGATCTCAGTCCAGTCCAGTGGAAA 314
|||||
QY 1868 gattgatacgttgccttggcgaacatgtatgttggaaagattcagacctgtctgtgc 1927
|||||
DB 315 GATTGATCAGCTGCTGTGGGAATGTGATTTGGAAAGATTTCACACCTGCTGGTGC 374
|||||
QY 1928 ggcactcaagcagatgcttggcctgctgctgctgctgctgctgctgctgctgct 1987
|||||
DB 375 GGCATCTCAAGCATGCTTTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 434
|||||
QY 1988 attccggagacacatgacctgctgctgctgctgctgctgctgctgctgctgct 2047
|||||
DB 435 ATTCCGGGGAGACCATGCTCGGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 494
|||||
QY 2048 tggatattgaagcagccttgggaagattgttggaaaggaagcag 2092
|||||
DB 495 TGAATCATGAAGCCACCTGGAAGATGTTTGGAAAGGAAGCAG 539
|||||

RESULT 25
AA243700 531 bp mRNA EST 07-MAR-1997
LOCUS AA243700 531 bp mRNA EST 07-MAR-1997
DEFINITION zrb8908.s1 Soares.NhHMPu.S1 Homo sapiens cDNA clone IMAGE:668606 3'
similar to SW:YK59.YEAST P36159 HYPOTHEETICAL.96.8 KD PROTEIN IN
SIS2-MTD1 INTERGENIC REGION.; mRNA sequence.
ACCESSION AA243700
VERSION AA243700.1 GI:1874492
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Hominidae: Homo.
REFERENCE 1 (bases 1 to 531)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
The Washu-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seg primer: -4iml3 fwd. RT from Amersham
High quality sequence stop: 466.
Location/Qualifiers

FEATURES

source

1. 531

Location/Qualifiers

/organism="Homo sapiens"
/db_xref="GDB:5562573"
/db_xref="taxon:9606"
/clone="IMAGE:668606"
/clone_1lb="Soares.NhHMPu.S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below): Vector: pTZ19-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBH, pregnant uterus
NbHPU, and fetal heart NbH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT

144 a 137 c 143 g 107 t

ORIGIN

Query Match 18.5%; Score 458; DB 4; Length 531;
Best Local Similarity 100.0%; Pred. No. 1.4e-230;
Matches 458; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 201 ccggagactggggcgccgctctacgtctctccgaattcaacggatctctcaactg 260
|||||
DB 74 CCGGGACTCGGGCCCGCGCTTACGCTTCTCCGAGTTCAACCGATCTCTTCAACTG 133
|||||
QY 261 tggagagcgcttgaagactcatgagagagcagaagttgaagttgctgcctggacaa 320
|||||
DB 134 TGGAGAAGCGCTTACGAGACTCATGACGAGACACAAATTAAAGTTCTGCTGGACAA 193
|||||
QY 321 catactctgacacgaatgactgcttaatgttggggcttgaagtgaatgattcttac 380
|||||
DB 194 CATATCTCTGACACGAATGCACTGTCTAATGTTGGGGGCTTAAAGTGAATGATTTTAC 253
|||||
QY 381 tttaaggaacccgggcttccaaagtgttactcttctggacctccaaactgaaata 440
|||||
DB 254 TTTTAAAGAAACCGGCTTCAAAAGTGTACTTCTGACCTCCCAACATGGAAAAATA 313
|||||
QY 441 ccttgaagcaatcaaatatttcttgctcatgtaagagatgaagctgctgctggcc 500
|||||
DB 314 CTTCGAAGCATCAAAATATTTTCTGCTCAATTGAAAGATGAATGCTGCTGCGGCC 373
|||||
QY 501 ccaactgtccccaagatacagagatgaacacatgacagttaccagattcccatatacag 560
|||||
DB 374 CCACTGTGCCCCAGAAATACAGAGATGAACACATGACAGTTTACAGATGCCCATACAG 433
|||||
QY 561 tgaacagagaggggaagcacaacatgagcagatcgaagaagccttcagcagct 620
|||||
DB 434 TGAACAAGAGAGGGGAAAGACCAACCATGGCAGAGTCCAGAAAGGCTCTCAGCAGGCT 493
|||||
QY 621 cagttcagagcagatcttcagactccagatgtaagaa 658
|||||
DB 494 CAGTCCAGAGCATCTCAGATCCGAGTGAATGAAA 531
|||||

RESULT 26
BG471513 611 bp mRNA EST 21-MAR-2001
LOCUS 602513052P1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4635685 5',
DEFINITION mRNA sequence.
ACCESSION BG471513 GI:13403788
VERSION BG471513.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 611)
NIH-MGC http://mgs.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM1393 row: 9 column: 14
High quality sequence stop: 590.
Location/Qualifiers
1. 611
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4635685"
/clone_1id="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 145 a 171 c 161 g 134 t
ORIGIN
Query Match 18.3%; Score 453; DB 154; Length 611;
Best Local Similarity 99.6%; Pred. No. 6.1e-228;
Matches 553; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 705 ggaacttccctggtgtagcttccatctgtaagcttaagaagaagaactctt 764
|||||
DB 19 GGACTCTTCCTGGTCGAGCTTTCATCTGTAGCTTCACTTAAAGAGAGAACTTCTT 78
QY 765 ggtgtctcaaaagagagatgaggccctcaggttgaggagacgtccatcgctcccatcat 824
|||||
DB 79 GGTCCTCAAAAGCAAGAGATGGGCTCCAGTTGGGACACTGCCATCGCTCCCATCAT 138
QY 825 tgcgtctgtaagaagcaggaagacatcatcatcatgtaagaagaagatttgctgaaga 884
|||||
DB 139 TGCTGCTGTCAAGAGAGGGAAGACATCATCATGAAAGAGAGATTGGTGAAGA 198
QY 885 gctgtgtaactctccagatctctggtgctcttggtgtgtagaatgtaagatgaag 944
|||||
DB 199 GCTGTGTAATCTCTCAGATCTCTGTCGTCTTGTGTGTAGTAATCTCCAGATGAAG 258
QY 945 cttaattcaaccatctgtgagaatgccaacttcaagaaggtacccaagaagaagcagatgc 1004
|||||
DB 259 CTTTCATTAAACCATCTGTGAGATGCCACTTTTCAGAGGTACCAAGAGAAAGGCGATGC 318
QY 1005 ccccgctgacctggtgtgtaacatgagccagcatctgtctgtgtaacagatgacca 1064
|||||
DB 319 CCCCGTGGCCTTGCTGTGATGACATGGCCCGAGCATCTGTGTTGTGAGACAGAGTACCA 378

QY 1065 gcaatgagatgagagatgttggtggtccctgacacccagcaacttggtctcgaatggaactgtgc 1124
|||||
DB 379 GCAGTGGATGGAGAGAGTCCGGGCTCGACACCCAGCACTTGCTCGAATGAGAACTGTGC 438
QY 1125 ctcaattacacactctgcagcaccacaagaattcaaacccagctcaacctcatcacccgga 1184
|||||
DB 439 CTCAGTTTCAACACCTTGCACGACCAAGATTTCAAAACCCAGCTCAACCTCATCCACCCGA 498
QY 1185 catctccctctgtctacacagttccgctgtaagaagaggccccacccctcagtgctc 1244
|||||
DB 499 CATCTTCCCTCCCTGCTCACACAGATTTCCGCTGTAGAGAGAGGCCCCACCTCATGTGCC 558
QY 1245 catgttcagggtga 1259
|||||
DB 559 CATGCTTCAGGGTGA 573
RESULT 27
A1200296
LOCUS A1200296 451 bp mRNA EST 14-OCT-1998
DEFINITION q186b12.x1 Soares_fetal_lung_NbHL19w Homo sapiens cDNA clone
IMAGE:1756895 3' similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8
KD PROTEIN IN SIS2-MTDL INTERGENIC REGION. ;, mRNA sequence.
ACCESSION A1200296
VERSION A1200296.1 GI:3752902
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 451)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400p from Gldco
High quality sequence stop: 442.
Location/Qualifiers
1. 451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1756895"
/clone_1id="Soares_fetal_lung_NbHL19w"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: p773D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
15'-TGTACCAATCTGAGTGGGAGCGCGCAATTTTCTTTTCTTTT-3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19w."
BASE COUNT 110 a 120 c 128 g 93 t
ORIGIN
Query Match 18.2%; Score 451; DB 17; Length 451;
Best Local Similarity 100.0%; Pred. No. 6.9e-227;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 90 gcggcgcgcaagagaccgcgtcgcgacactgcgcagcagagaagcgcgagcgtcggg 149
|||||
DB 1 GCGGCGCGCAAGAGACCGGCTGGCGCACCTGCGCACGAGAGAGAGACCGGCGCGG 60

Oy	150	gtgctccggcgagccccaaaccgcgtgtaacctgcgaagtgtgtgcagcgcgtagtgcgggactc	209
Db	61	gtgctccggcgagccccaaacacgcgtgtaacctgcgaagtgtgtgcagcgcgtagtgcgggactc	120
Oy	210	ggcgcgcgctctcaatgcgtcttccttcgcgaattcaaacggtatctcttcaactgtgtgagaag	269
Db	121	ggcgccccgcgcgtcaacgcttcttcgccgattcaacccggtatctcttcaactgtgtgagaag	180
Oy	270	cgttcaggagactcatgatgcagagacaagtaaagaattgctgcgccttggaagaatatctct	329
Db	181	cgttcaggagactcatgatgcagagaccaaatgrraaagttgctgcgccttggaacaatactct	240
Oy	330	gacacgaatgcactggtgtctaagtgtgggggttaagtgaatgatcttactttaaga	389
Db	241	gacacgaatgcactggtgtctaagtgtgggggttaagtgaatgatcttactttaaga	300
Oy	390	aaccgggttccaagaatgtgtactcttcttggaactccaactgtgaaaaataactcgaagc	449
Db	301	aaccgggttccaagaatgtgtactcttcttggaactccaactgtgaaaaataactcgaagc	360
Oy	450	aatcaaatattcttcgtccatctgaagaagaaatgaactgctgtgtgcgcccactctgc	509
Db	361	aatcaaatattcttctgtccatctgaagaagaaatgaactgctgtgtgcgcccactctgc	420
Oy	510	cccagaatacgaagatgaaacatgcacgtt	540
Db	421	cccagaatacgaagatgaaacatgcacgtt	451
RESULT_28			
LOCUS	BE867512	812 bp	mRNA
DEFINITION	601443010F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847226 5',	EST	20-OCT-2000
ACCESSION	BE867512		
VERSION	BE867512.1	GI:10316288	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 812)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgaps-email.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov plate: LLAM561 row: c column: 03 High quality sequence stop: 686. Location/Qualifiers 1..812 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3847226" /clone_lib="NIH_MGC_65" /tissue_type="adenocarcinoma" /lab_host="DH10B (phage-resistant)" /note="Organ: colon; Vector: pCMV-Sport6; Site: 1; Note: Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT Average insert size 1.8 kb. Library constructed by Life Technologies."		
FEATURES			
SOURCE			
BASE COUNT	200 a	216 c	220 g 176 t
ORIGIN			

Query Match	18.1%	Score 450	DB 141	Length 812
Best Local Similarity	99.8%	Pred. No. 2.4e-226		
Matches 570	Conservative 0	Mismatches 0	Indels 1	Gaps 1
QY 724	gcttcatctgttaacttacttaagaagaagaactcttgctgtctcaagaagaag	783		
Db 1	gcttttcattctgtaaccttactttaaagagagaaactcttggtctcaagaaagag	60		
QY 784	atggagctcccaagtttggaacagctgcacatgcctccatcatctgtctgtcaagaagg	843		
Db 61	atggagctctccagtttggaacagctgcacatgcctccatcatctgtctgtcaagaagg	120		
QY 844	aaaacatcatcactgaagaagaagaatttgcttgtaagaagctgtgtacccctccaat	903		
Db 121	aaaacatcatcactgaagaagaagaatttgcttgtaagaagctgtgtacccctccaat	180		
QY 904	cctgtgtctgtcttctgtgtgtgaagaagtcacagaatgaagaatcattcaaccatcgt	963		
Db 181	cctgtgtctgtcttctgtgtgtgaagaagtcacagaatgaagaatcattcaaccatcgt	239		
QY 964	gagaaatgcacacttcagaagtaaccaagaagaagcagatgcctccgttgctgtgtgt	1023		
Db 240	gagaaatgcacacttcagaagtaaccaagaagaagcagatgcctccgttgctgtgtgt	299		
QY 1024	caacatgcccccaagcactgtgtgtgttgtaagaagcagatgcacagaagtgatagaagttc	1083		
Db 300	caacatgcccccaagcactgtgtgtgttgtaagaagcagatgcacagaagtgatagaagttc	359		
QY 1084	ggagctgacacacacacactgtgtgtgtgtgaatagaactgtgctcagttcaaaccttcgc	1143		
Db 360	ggagctgacacacacacactgtgtgtgtgtgaatagaactgtgctcagttcaaaccttcgc	419		
QY 1144	agccacaagatcaaac	1203		
Db 420	agccacaagatcaaac	479		
QY 1204	agtttcgcctgttaagaagaagaagggccccacacacacacacacacacacacacacacac	1263		
Db 480	agtttcgcctgttaagaagaagaagggccccacacacacacacacacacacacacacacac	539		
QY 1264	ctccctcaagatcac	1294		
Db 540	ctccctcaagatcac	570		

RESULT 29

BF433563

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BF433563 498 bp mRNA EST 30-MAR-2001

7455b07.x1 NCI-CGAP_Lu24 Homo sapiens CDNA clone IMAGE:3702108 3'

Similar to TR:Q9V5J4 Q9V5J4 CG3298 PROTEIN. ; mRNA sequence.

BF433563

BF433563.1 GI:11445735

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 498)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -400P from Gibco

High quality sequence stop: 459.

FEATURES Location/Qualifiers

SOURCE

1..498

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3702108"

/clone_1id="NCI CGAP Lu24"

/tissue_type="carcinoid"

/lab_host="DH10B"

/note="Organ: Lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI CGAP Lu25 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 132 a 130 c 131 g 105 t

ORIGIN

Query Match 18.0%; Score 447; DB 148; Length 498;

Best Local Similarity 99.8%; Pred. No. 9.1e-225;

Matches 497; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 142 ccgtcgggtgtctccggcgcccaacacgttactctgcaggtgtgtgcagcggtagc 201

Db 1 cccgtcgggtgtctccggcgcccaacacgttactctgcaggtgtgtgcagcggtatc 60

QY 202 cggagctcggcgccgctcctacgtcttcctccaggttccaccggtatcttcaactgt 261

Db 61 cggagctcggcgccgctcctacgtcttcctccaggttccaccggtatcttcaactgt 120

QY 262 ggaagaagcggttcagagatcatgagagagacaaagttaagttgtctgcctggacac 321

Db 121 ggaagaagcggttcagagatcatgagagagacaaagttaagttgtctgcctggacac 180

QY 322 atattctcgaagagatgactgttctaaagtgtggggccttaagtggagatatttctct 381

Db 181 atattctcgaagagatgactgttctaaagtgtggggccttaagtggagatatttctct 240

QY 382 ttaagaagaacggggtctccaaggtgttactttcttgaccctccacaacttgaaaaatac 441

Db 241 ttaagaagaacggggtctccaaggtgttactttcttgaccctccacaacttgaaaaatac 300

QY 442 ctcgaaagcaatcaaatatttctgtgtccattgaaaggaatagaactgtgtgtgcggcc 501

Db 301 ctcgaaagcaatcaaatatttctgtgtccattgaaaggaatagaactgtgtgtgcggcc 360

QY 502 cactctgcccagaatacagagatgaaacatgacagtttaccagatcccatcacagt 561

Db 361 cactctgcccagaatacagagatgaaacatgacagtttaccagatcccatcacagt 420

QY 562 gaacagaagagagaaacacacaaactgagcagagttccagaagaagccctcgaacaggtc 621

Db 421 gaacagaagagagaaacacacaaactgagcagagttccagaagaagccctcgaacaggtc 480

QY 622 agtccagagcagatctca 639

Db 481 agtccagagcagatctca 498

RESULT 30

BE260626

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Human.

Homo sapiens

REFERENCE 1 (bases 1 to 938)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LNCMI9 row: 0 column: 04

High quality sequence stop: 621.

FEATURES

source

1..938

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3161691"

/clone_1id="NIH-MGC_19"

/tissue_type="neuroblastoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: POTB7; Site: 1; Site: 2; EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 190 a 263 c 283 g 202 t

ORIGIN

Query Match 18.0%; Score 447; DB 165; Length 938;

Best Local Similarity 100.0%; Pred. No. 9.4e-225;

Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 916 ttgtgtgtgtatgaatgttcagatgaagaacttcaatccatctgtgagaatgacac 975

Db 135 ttgtgtgtgtatgaatgttcagatgaagaacttcaatccatctgtgagaatgacac 194

QY 976 ttccagaggtaccaaaggaagcagatgtcccccgtgtgcttgggttccaatgtgccca 1035

Db 195 ttccagaggtaccaaaggaagcagatgtcccccgtgtgcttgggttccaatgtgccca 254

QY 1036 gcatctgtgtgtgtgagacagcaggtaccagcaggtgagagaggtttgggacctgacac 1095

Db 255 gcatctgtgtgtgtgagacagcaggtaccagcaggtgagagaggtttgggacctgacac 314

QY 1096 cagcactgtgtccttgaatgaagaactgtgctcctaagttcaacaccttcgcagccacaagatt 1155

Db 315 cagcactgtgtccttgaatgaagaactgtgctcctaagttcaacaccttcgcagccacaagatt 374

QY 1136 caaacccagctcaacctatccaccggagacatcttccctgtgtccacagtttccgtgt 1215

Db 375 caaacccagctcaacctatccaccggagacatcttccctgtgtccacagtttccgtgt 434

QY 1216 aagaagagagagcccccacacctgaagtgtgccatgtgttaagggtgaatgctctcctaagttac 1275

Db 435 aagaagagagagcccccacacctgaagtgtgccatgtgttaagggtgaatgctctcctaagttac 494

QY 1276 cagctccgttcccaaggaaggtgtgcaagagagatgcatattacttgaatccttgaagaa 1335

Db 495 cagctccgttcccaaggaaggtgtgcaagagagatgcatattacttgaatccttgaagaa 554

QY 1336 ttcatagttgagggcgtcagcttccc 1362

Db 555 ttcatagttgagggcgtcagcttccc 581

[illegible][illegible]

Db 168 GACAGCAGTACACGACGATGAGAGGATTTGGGCTTGACACCCAGCAGCTGGTCTG 227

Qy 1111 aatgagaactgtgctcagttcacaacttgcagagccaaagtattcaaccagctcaac 1170

Db 228 AATGGAATCTGGCTTCAGTTCACAACTTCGACACCACAAAGATTCAACCCAGCTCAAC 287

Qy 1171 ctcatcacccggagacatctcccccgtctccacagttccgctgtaagaagaagggccc 1230

Db 288 CTCATTCACACCGGACATCTTCCCTGCTCACAAGATTCCCTGTGAAGAAGAGGGCCCC 347

Qy 1231 accctcagtggtccatgtgtcaggglyaatgctcctcaagtaccagctccgtccagg 1290

Db 348 ACCCTCAAGTGTGCCATGTTCAGGGTGAATGCCCTCTCAAGTACACAGCTCCGTCCAGG 407

Qy 1291 agggagtgagcagagagatgcatattactctgcaatcctctgagggaattcatgttgagcg 1350

Db 408 AGGAGATGGCAGAGGAGGATCATTTACTTCAATCTTGAGGAATTCATGATTGAGCGC 467

Qy 1351 ctgcagcttcccaacttcacagacagcgtgcagagatcacaggaagagtgccagagagc 1410

Db 468 CTGCAAGTCTCCCACTTCACAGAGAGCTGCAAGAGTACAGAGAGAGTCCGACAGACGCG 527

Qy 1411 ccagcccccagagagaa 1428

Db 528 CCAGCCCGCAGAGAAA 545

RESULT 33

AA455121 493 bp mRNA EST 06-JUN-1997

LOCUS z78c04.s1 Soares ovary tumor NBHOT Homo sapiens cDNA clone

DEFINITION IMAGE:809862.3', similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8

ACCESSION AA455121 GI:2177897

VERSION AA455121.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 493)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, J., Waterston, R. and Wilson, R.

TITLE WashU-Merck EST Project 1997

JOURNAL Unpublished (1997)

COMMENT Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Contact: Wilson RK

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

Seq primer: -41ml3 fwd. ET from Amersham

High quality sequence stop: 447.

Location/Qualifiers

1. 493

/organism="Homo sapiens"

/db_xref="GDB:6039680"

/db_xref="taxon:9606"

/clone="IMAGE:809862"

/clone_idb="Soares ovary tumor NBHOT"

/sex="Female"

/tissue_type="ovarian tumor"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: ovary; Vector: pTR73D (Pharmacia) with a

modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGGCTTTTCTTTTCTTTT 3']

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pTR73 vector

(Pharmacia). Library constructed by Bento Soares and

M. Fatima Bernaldo.

BASE COUNT 131 a 128 c 130 g 104 t

ORIGIN

Query Match 17.6%; Score 436; DB 7; Length 493;

Best Local Similarity 100.0%; Pred. No. 6e-219;

Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 202 cggagactggggcgccgagcttactgcttcccgagttcaaccggatcttcaactgt 261

Db 58 CGGAGCTGGGGCGCGCCTCTACGCTTCTCCGAGTTCAACGGATCTCTCAACTGT 117

Qy 262 ggaagaagcgttcaagaagctcatgcagagacacaagttaaaggttgcgcctgacacac 321

Db 118 GGAGAAAGCGTTCAAGAGACTCATGCAGAGACACAGTTAAAGTGTGCTCGCTGACACAC 177

Qy 322 atattcctgacaagaaatgcaactgtctaatgttggggccttaagtgaatgatttact 381

Db 178 ATATTCCTGACACAGAAATGCACTGCTAATGTTGGGGCTTAAAGTGAATGATTCTTACT 237

Qy 382 ttaaggaagaccgggcttccaaagtgttacttcttggaactcccaactggaanaatac 441

Db 238 TTAAGGAAGACCGGGCTTCCAAAGTGTACTTCTTGAGACCTCCACCACTGGAATAATAC 297

Qy 442 ctcaagaacatcaaatatttctgttcacatgaaagaatagaactgtgtgcggcc 501

Db 298 CTCGAAGCAATCAAAATTTTCTGTGTCATTTGAAGAATAGACTGTGCTGGGGCC 357

Qy 502 cactctgcccagaatagcagagatgaaacatgacagtttaccagatcccatcacagt 561

Db 358 CACTCTGCCCCAGAAATAGAGAGATGAACCAATGACAGTTTACCAATCCCATACACAGT 417

Qy 562 gaacagagagaggggaaacacacacacatgagagatcccaagaaggccctcagagctc 621

Db 418 GAACAGAGAGAGGGGAAACCAACCATGCGAGAGTCCAGAAAGGCTCTCAGCAGGCTC 477

Qy 622 agtccagagcagatct 637

Db 478 AGTCCAGAGCAATCTT 493

RESULT 34

A1468143 421 bp mRNA EST 30-MAR-1999

LOCUS A1468143

DEFINITION tf92905.x1 NCI-CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2106776.3'

Similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8 KD PROTEIN IN

S1S2-MTD1 INTERGENIC REGION.; mRNA sequence.

ACCESSION A1468143

VERSION A1468143.1 GI:4330233

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 421)

AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project

(CGAP/BTGAAP), Tumor Gene Index

Unpublished (1998)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,

Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bernaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Db 61 TGTCTCCGGCGCCCAACACCGGTACCTGACAGGTGTTGACGCGGAGCCGGACTCG 120
QY 211 ggccgcgcgcctcagctcttcctcagagttcaaccggtatctcttcaactcgggagaagc 270
Db 121 GGGCCCGCGCTCTACGCTTCTCCGAGTTCAACCGGTATCTTCACTGTGGAGAAAGC 180
QY 271 gttcaagactcagcaggaagacaaagttaaaggttgcgcctcgggaacatacttcctg 330
Db 181 GTTAGAGACACTAGCAGAGACACAGTTAAAGTTGCTCCCGTCGACAACTATTCCTG 240
QY 331 acacgaatgcacgtgtcctaattgttgggggcttaagtgaatgattcttacttaagaa 390
Db 241 ACAGCAATGACTGCTCAATGTTGGGGGCTTAAGTGAATGATTTCTTAAAGGAA 300
QY 391 accgggtctcacaagtgtgtactcttcggacctccaacacgggaaaaatctctgaaga 450
Db 301 ACCGGGCTCCAAAGTGTGTTCTTGACCTCCACAACTGAAATAATCTCCAAACA 360
QY 451 atcaaatattcttgcgtcattgaaagaatagaactgctgtgcggcccaactcgc 510
Db 361 ATCAAAATATTTTCTGCTGCTTGAAGAGATAGAACTGGCTGTGCGCCCACTCTGCC 420
QY 511 ccagaatacagcagatgaacacatgacagttaccagatcccatatacagatg 562
Db 421 CCAGAAATACGAGATGAACCATGACAGTTTACAGATCCCATACACAGTG 472

RESULT 36
LOCUS BG327066 452 bp mRNA EST 27-FEB-2001
DEFINITION 602426274F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564170 5',
mRNA sequence.
ACCESSION BG327066
VERSION BG327066.1 GI:13133503
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 452)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabds-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1279 row: c column: 19
High quality sequence stop: 452.
Location/Qualifiers

FEATURES
source
1..452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4564170"
/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Kidney; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 93 a 132 c 144 g 83 t

Query Match 16.4%; Score 406; DB 152; Length 452;
Best Local Similarity 100.0%; Pred. No. 4.4e-203;
Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 cgagacgacacattgtcgcagagagacacacatactcaagacacccgcgcgcgagcgcc 95
Db 47 CGGAGCGACCATGTGTGCGAGGAGCAGCATATTCGAGGACACCGCCCGCGAGGCGCC 106
QY 96 gcgcgaagaccgcgtgcggacctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 155
Db 107 GCGCAAGGACCCGCGTGCAGCAGCTGCAGCAGCGAGAGAGGCGGAGCCGCGGCTGCTC 166
QY 156 cggcgggcccaaacaccggtatccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 215
Db 167 CGCGGCGCCCAACACCGGTACTGTACAGTGTGTGACCGGAGTACCGGAGCTCGGCGC 226
QY 216 cgcgctcagctctctcgc 275
Db 227 GCGGCTGACGCTTCTTCTCGAGTTCAACCGGATCTCTTCAACTGTGAGAAAGCGTTCA 286
QY 276 gagactcagcagagacacaaagttaaaggtgtcgcgcgcgcgcgcgcgcgcgcgcgcgc 335
Db 287 GAGACTCATGACGAGGACACAAAGTTAAAGTTGCTGCGCTGAGCAACATATTCCTGACAG 346
QY 336 aatgcacgtgtcctaattgttgggggcttaagtgaatgattcttacttaaggaacccg 395
Db 347 AATGCACGTGTCTAATGTTGGGCGCTTAAGTGAATGATTTCTTAAAGGAAACCGG 406
QY 396 gcttccaaagtgtgactcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 441
Db 407 GCTTCCAAAGTGTGACTTCTTGACCTCCACACTGGAATAATATAC 452

RESULT 37
LOCUS BF309436 457 bp mRNA EST 21-NOV-2000
DEFINITION 601892128F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4137654 5',
mRNA sequence.
ACCESSION BF309436
VERSION BF309436.1 GI:11256824
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 457)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
Plate: LNCM1045 row: h column: 07
High quality sequence stop: 457.
Location/Qualifiers

FEATURES
source
1..457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4137654"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pORF7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by

QY 1860 agtggaaagattgctcagttcgtcgttgcggaactgtgatt 1900
|||||
Db 480 AGTGAAGAATGATGATCGCTGTTGCCGAACATGTGATT 520

RESULT 41
BE747163
LOCUS BE747163 992 bp mRNA EST 15-SEP-2000
DEFINITION 601577254F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838237 5',
mRNA sequence.
ACCESSION BE747163
VERSION BE747163
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 992)
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LCM524 row: 1 column: 14
High quality sequence stop: 781.
Location/Qualifiers
1. .992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3838237"
/clone_1ib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOT7; Site:1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(6). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 225 a 287 c 313 g 167 t

ORIGIN

Query Match 16.1%; Score 400; DB 139; Length 992;
Best Local Similarity 99.7%; Pred. No. 6.9e-200;
Matches 570; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1082 ttgggctcgcaccccgacctgtgctcctgaatgagaactgtgctcagttcacaccttc 1141
|||||
Db 90 TTGGGCTGACACCCAGCACTTGCTCTGAATGAGAACTGTGCTCAGTTACACACCTTC 149

QY 1142 gtagcccaagattcaaacccagctcaactatcaccacccggagacattcccccgtctca 1201
|||||
Db 150 GCAGCCACAGATTCAAAACCCACCTCAACCTCATCCACCCGGACATCTTCCCTGCTCA 209

QY 1202 ccagtttcgctgtaagaaggagcccccacccctcagttgtgccatggttcagggtgaat 1261
|||||
Db 210 CCAAGTTCCGCTTAAGAGAGAGGCCCCACCTCTCACTGTGCCCCATGTTCAAGGTAAT 269

QY 1262 gctcctcagtaacagctcctcccaaggaggagtggcagaaggatgcatattactt 1321
|||||
Db 270 GCCTCTCTCAAGTACAGCTCCGTCGACGAGGAGGAGTGCAGAGGATGTCATTATTACTT 329

QY 1322 gcaatcctaaggaaatcatagttaggcgtgtgagcttcccaactccagcagaagcgtgc 1381
|||||

Db 330 GCAATCTGTAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCACTTCACGACGCGTC 389
QY 1382 agaggtacagagagagtgccgacgagcgccacccacgacagagaaagagtaacac 1441
|||||
Db 390 AGGAGTACAGAGAGAGAGTGGCCAGAGCGG -CCAGCCCAACAGAGAAAGATGATGATACC 448

QY 1442 cagaatcatcttcctcttgagacagaggtctgcacatcccatgaagattcgaaatgctag 1501
|||||
Db 449 CAGAAATCATCTCTCTTGGAACAGAGGTCTGCATCCGATGAGATTCGAATATGTCAGTG 508

QY 1502 ccgacctgtcaacataagaccgccgacacgctctcgtcctacagtgatgtgtgaggacat 1561
|||||
Db 509 CCACACTTGTCAACATTAAGCCCGACACGTCCTTGCTACTGAGCTGTGTGAGGCGCAT 568

QY 1562 ttggcagctgtgcgcgtcattacagagaccaggtggagcaggttctgtggcaccctgctg 1621
|||||
Db 569 TTGGGCAAGCTGTGGCGTCATTTACGAGACACAGGTGGAGAGGCTCTGGACCTGGCTG 628

QY 1622 ctgtgtttgtgtccacctgacgcagatcac 1653
|||||
Db 629 CTGTGTTGTGTGTCACCTGCACGTCACGACAGATCAC 660

RESULT 42
BE386924
LOCUS BE386924 627 bp mRNA EST 21-JUL-2000
DEFINITION 601274815F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3615939 5',
mRNA sequence.
ACCESSION BE386924
VERSION BE386924.1 GI:9352387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 627)
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LCM280 row: n column: 04
High quality sequence start: 18
High quality sequence stop: 611.
Location/Qualifiers
1. 627
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3615939"
/clone_1ib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOT7; Site:1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(6). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 147 a 175 c 199 g 106 t

ORIGIN

Query Match 15.6%; Score 387; DB 167; Length 627;
Best Local Similarity 100.0%; Pred. No. 5e-193;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2095 gaaagacacacagcacaacgtcccaagccatcagctgtggagatgcgatatgaacgcgag 2154
|||||
Db 16 GAAAGACACACAGCACAACGTCCTCCAGCCATCACGCTGGGGATCGGATGAACCGGAG 75
QY 2155 ttcattatcgttaaacattcagccagcgcatactgccaaggtccctctcagccccaac 2214
|||||
Db 76 TTCAATTATGCTTAACACTTTCAGCCGCTATGCCAAGTCTCCCTCTCTCAGCCCAAC 135
QY 2215 ttcagcgaagaatgtggagctgtgcttgcacacatgaagctgtctgttggagacttcca 2274
|||||
Db 136 TTCACGCGAAGAAAGTGGAGTTTCCTTTGACACATGAAGGTCTGCTTTGGAGACTTTC 195
QY 2275 acaagcgaagctgtatcccccactgaaagccctgttgcctggaggaatcgaagaagt 2334
|||||
Db 196 ACAATGCCCAAGCTATATTCCTCCACTGAAAGCCCTGTTTGGGACATCGAGAGATG 255
QY 2335 gagagcagcaggaagaacgcgagagctgcgcaggtgcgcgcctccctgcacaggaag 2394
|||||
Db 256 GAGGAGCCGACGAGGAGAGAGGAGCTGCGGCGAGTGGCGGCGCCCTCTCTCCAGGAG 315
QY 2395 ctggcagcgcgcctgtggagatgtggagagctcagcagaagcgcgcacacagagaagcca 2454
|||||
Db 316 CTGGCAGCGCGCTGAGAGATGGGAGCCTCAGCAGAAAGCGGCGCCACACAGAGAGCCA 375
QY 2455 cagggccaagaagtctagagccagtgta 2481
|||||
Db 376 CAGGCCAAGAAAGTCAAGACCCAGTGA 402

RESULT 43

LOCUS BE250309 574 bp mRNA EST 13-JUL-2000
DEFINITION 600943455F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960077 5',
mRNA sequence.
ACCESSION BE250309
VERSION BE250309.1 GI:9120418
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 574)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov
Plate: LCM52, row: n, column: 14
High quality sequence stop: 571.

FEATURES

source

Location/Qualifiers
1..574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2960077"
/clone_1ib="NIH_MGC_17"
/tissue_type="thadomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pORF7; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(6). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

124 a 158 c 167 g 119 t 6 others

ORIGIN

Query Match 15.5%; Score 384; DB 165; Length 574;
Best Local Similarity 99.8%; Pred. No. 1.9e-19;
Matches 434; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1574 gccgtcattacggaacacaggtgtgacaggtgtcgtggacacctgtgtgtgtgtgt 1633
|||||
Db 1 GCCGTATTACCGAACCAGGTGACAGGTCCTGGACCTGGCTCTGTTGTGT 60
QY 1634 cccacctgaacgcagatcaaccacagggcttgcagatattctgtcgaagaagacg 1693
|||||
Db 61 CCCACTGACACCAATCATCACACAGGGCTTGCCAGATATCTTGCTGCAGAGAACCG 120
QY 1694 ccttgcatttgggaagccgcttccaccttgcctgtgtgtgtgtgtgtgtgtgt 1753
|||||
Db 121 CTTTGGCATCTTTGGGAAAGCCGCTTCAACCTTTGCTGTGTGTGTGTGTGTGTGT 180
QY 1754 aagctgtgtcctcaggaagtaaccacaggtgcgaagaggtcctgcacacatcag 1813
|||||
Db 181 AAGCTGTGCTCCAGACATACCAACCAAGTGCAGAGAGTCTTGCACACATCATATGA 240
QY 1814 ttcctgcgaatgcctcaggaagggcgtgaagatccagtcctcagtggaagattga 1873
|||||
Db 241 TTCCGTGCCAATAGCTTTCAGAGAGGGCTGAGATCTCCAGTCTTCAGTGAAGATGA 300
QY 1874 tcaattcgtgttgcgaacatgtgatttggaaagattcagaacctgtgtgtgtgtgt 1933
|||||
Db 301 TCAGTTCCGCTTTGGGAAAGTGTGATTGGAAAGATTTCACACCTGTGTGTGTGTGTGT 360
QY 1934 gaaagcagctgt 1993
|||||
Db 361 GCAAGCATGCGCTTGTGCGTGTGTCGTCGACACCTGTGCGAAGGTCTATTTCG 420
QY 1994 gggacacatggcct 2008
|||||
Db 421 GGGACACCATGCTCT 435

RESULT 44

LOCUS BF732987 397 bp mRNA EST 30-MAR-2001
DEFINITION nae07b01.x1 NCI-CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3434448 3',
similar to TR:Q9V5J4 Q9V5J4 G33298 PROTEIN. ;, mRNA sequence.
ACCESSION BF732987
VERSION BF732987.1 GI:12058062
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 397)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN, send email to:
info@image.lnl.gov
Seq primer: -400p from Gibco
High quality sequence stop: 354.

FEATURES

source

Location/Qualifiers
1..397
/organism="Homo sapiens"
/db_xref="taxon:9606"

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 20:05:27 ; Search time 1393.91 Seconds
(without alignments)
11107.759 Million cell updates/sec

Title: SEQLINS_COPY_1140_2140
Perfect score: 1001
Sequence: 1 tcgcagccacacagattcaaa.....aagccatcagcgtgggatg 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 773874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rtd: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rtd: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

Insertion
of seq 10 NO1
std
search of
1000 residues
bracketing
insertion
site

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	989	98.8	2908	89	AF304370	AF304370 Homo sapi
2	987.4	98.6	2997	91	BC001939	BC001939 Homo sapi
3	987.4	98.6	3006	91	BC004158	BC004158 Homo sapi
4	973	97.2	2908	89	AF308698	AF308698 Pan trogl
5	971.4	97.0	2976	89	AK001392	AK001392 Homo sapi
6	965	96.4	2893	89	AF308694	AF308694 Gorilla g
7	728.8	72.8	2712	94	AF308696	AF308696 Mus muscu
8	137	13.7	118788	85	AC005277	AC005277 Homo sapi


```

QY 481 tgcgtcttctgtccacccctgagcaagatcacacacagcgttcccaatctctgc 540
      |||||||
Db 1620 TGCCTGTTTGTGTCCACCT -GCACCGAGATCCACACGGGCTTGCAAGTATCTTGC 1678
QY 541 tgcagagagaagcgccttgcacatcttgggaaagcgcctacccttgcctgtg 600
      |||||||
Db 1679 TGCAGAGAGAAGCGGCTTGSCATCTTGGGAAGCGGCTTCCACCTTGTGTGTG 1738
QY 601 cccccaacagctcaaacctgtgctccagcaagtaaccaacagtcgcaagagtcctgc 660
      |||||||
Db 1739 CCCCCAACCGAGCTCAAGGCTGGCTCCAGAGTACCAACACAGTGCAGAGAGGCTCCTGC 1798
QY 661 accacatcatatgatctctcccaaatgacctcaaggaagagcgcagatctccagctcgt 720
      |||||||
Db 1799 ACCCATCATATATATCTTCCGCCAATGCTTTCAGGAAGGGGCTGAGATCTCCAGTCTG 1858
QY 721 caagtgaagaatgatcatagctcgtctgtgcgaacatgtgatttggaaagattcacacct 780
      |||||||
Db 1859 CAGTGAAGAAGTATGATCATGTTCCGCTGTTGGGACATGATGATTGGGAAGATTTCAGACCT 1918
QY 781 gctcgtgtgcgcactgcagacatgcgttctgtgcgtgtgcacacacctgtgcctga 840
      |||||||
Db 1919 GTCGTGTCGGGACACTGCAGAGATGCTGTTGGCTGTGCTGCTGCACACCTCTGCTGGA 1978
QY 841 aagtggtctatctccgggagcaacacatgcccctgcagagctctgtctcgatgggaaagatg 900
      |||||||
Db 1979 AAGTGTCTATCTCCGGGACACCATGCTGCGAGGCTCTGTGTCGGGATGGGAAAGATG 2038
QY 901 ccacccctccgatatacatgaagccaccccttggaagattgttgaaagagaagcaatgtgaa 960
      |||||||
Db 2039 CCACCCCTCGATCATGATGAAGCCACCTCGAAGATGTTTGAAGAGAACACTGTGAAA 2098
QY 961 agacacacagacaaagtcaccaagcatcagcgtgggagatg 1001
      |||||||
Db 2099 AGACACACAGCACAACTCCCAAGCCTACAGCTGAGGGATG 2139

```

```

FEATURES
Source
1. 2997
/organism="Homo sapiens"
/db_xref="LocusID:60528"
/db_xref="taxon:9606"
/clone="MGC:4102 IMAGE:2820640"
/tissue_type="Lung, small cell carcinoma"
/clone_id="NIH_MGC_7"
/lab_host="DH10B-R"
/note="Vector: pOT79"
61..2541
/codon_start=1
/product="putative prostate cancer susceptibility protein"
/protein_id="AAH01939.1"
/db_xref="GI:12804973"
/translation="MMAALCSLNSAGRTMSGRTISQAPARREPRKDPRLRLTRE
KRGPSGSGPNTVYLQVVAAGSRDGAALYRSEFRRLYFNGEGYQRLMOEHLKY
ARLDNFIETPMHMSNWGLSMILITLKEGLPFCVLSGPQLKEAIKIFSGPLKG
IELAVRPHSAPREYEDETMTVOIPIHSEORGRHOMQSPERPLSLSPRSSSEIN
ENBPILPHGVSGRGRVDSLLVAATFLKLRKGNPLYLAKEMGLPVTAAIAPITA
AYDKSITHEGREILAEELCTPPDEAAVYVECPDESTIOPTCEATQROGRKD
APALVYHMAPASVLYDSRYQOMMERGRPTQHLVLNENCAVHNLSHR1QQLNLI
HPDIFLTFSEKRCKEGPTLSVPWQOESQPELKYOLRRERMOBDAITTCNPEEYVA
LOLPNEQTSQVEYRARSADGPAPAEKRSQPELITFETGSAIPKINVAATVNISSP
DTSLDDCGGTGOLCRHYGDQVDVRLTAAVFEVHLADHTGTGPISTLDERAL
ASLGRPLHPLVVAAPNOLKAMLDQYHNOCEVLAHISMPAKLOEAEFISPAVERL
ISSILRTCDLEEFOTCLVRHCKAFGALVHTSGWKVYSGDTMPCALVRMGKDAL
LIHATLEDDLEEAEEVETKSTTSQALISVGMRNAEETIMLHFSQRYAKVPLSPNS
EKYVAVDHMKVCFGDEPTMPKLIPLKALFAGIDIBEMERREKRELOVRAALLSRE
LAGLEGCEPQOKRAHTEEPQAKVRQ"
BASE COUNT 740 a 806 c 853 g 598 t
ORIGIN

```

```

Query Match 98.6%; Score 987.4; DB 91; Length 2997;
Best Local Similarity 99.8%; Pred. No. 2.8e-261;
Matches 999; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 tgcagagccacaagatccaacccagctcaacctacatccaccggagacatctccctcgt 60
      |||||||
Db 1200 TCGCAGCCCAAGATTCAAAACCCAGCTCAACCTCATCCACCGGACATCTCCCTCTCT 1259
QY 61 caccagcttcgcgtgtaagaagagggcccaacctcagctgtgcccattgtaagggtga 120
      |||||||
Db 1260 CACGAGTTTCCGCTGAAGAGAGGGCCCAACCTCATGTTGCCCAAGTTCAGGGGTGA 1319
QY 121 atgcctcccaagtaaccagctccgtcccaagagagagtgagagggagatgcatattac 180
      |||||||
Db 1320 ATGCTCTCTCAAGTACACAGCTCCGTCACAGAGAGGAGGAGGAGATGCAATTATTAC 1379
QY 181 ttgcaatcttgaggaatcatagttgagggcgtcagcttcccaactccagcagagcgt 240
      |||||||
Db 1380 TTGCAATCTTGAGGAATTCATAGTTGAGGCGCTCAGCTTCCCAATCTCCAGCAGAGCGT 1439
QY 241 gcaagagtaacagagagtgtagcagagcggcccaagcccaagaaaagaagtaagta 300
      |||||||
Db 1440 GCAGGAGTACAGGAGAGATGGCAGAGGCGCCACGCCCAAGAAAGAAAGTCAAGTA 1499
QY 301 cccagaaatcatcttcccttgaaacagggctgcacatcccgatgaagattcgaatgtcag 360
      |||||||
Db 1500 CCGAAGAAATCATCTTCTTGGAACAGAGGTCTGCCATCCCGATGAAGATTCCAATGTGAG 1559
QY 361 tgcacactgttcaacataaagcccgacaagctctcgtactagactgtgtgagggcac 420
      |||||||
Db 1560 TGCACACACTTGTCAAACTAAACCCCGACACGCTCTGCTACTGAGACGTGTGAGGGGAC 1619
QY 421 attggagcagctgtgcgctcatcagagaccaggttgagacgggttcctcggcacctggc 480

```

Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 9 Row: 1 Column: 17.

Location/Qualifiers

```
|||||
Db 1620 GTTTGGGAGCTGTGCGCTCATTTACGAGAACCGAGGTGACAGAGGTCCTGGGACCCCTGGC 1679
Qy 481 tgcctgtgtgtgtccacccttgcacgcagatcacacacggcgttgccaatctctgc 540
Db 1680 Tcctgtgtgtgtgtccacccttgcacgcagatcacacacggcgttgccaatctctgc 1738
Qy 541 tgcagagagaagcgcgccttgcacatcttgcagaaagcgcgttcacaccttgcctgtgtgc 600
Db 1739 TGCAGAGAGAACGCGCGCTTGGCATCTTTGGGAAAGCGGCTTACCCCTTGTGCTGTG 1798
Qy 601 ccccaacacagctcaaaagcctgtgtctcagcagtaaccacaacagctgcagaggtctgc 660
Db 1799 CCCCACACAGCTCAAAAGCTGCTGCACACAGTACCACACAGGCTCCAGAGGTCCTGC 1858
Qy 661 accacatcagtaagttctctgcacaaatgccttcagaaagggctagatctccatctcctg 720
Db 1859 ACCACATCATGTATGATTCCTGCGCAATGCTTCAGAGAAAGGCTGAGATCTCCAGTCTG 1918
Qy 721 cagtgaaagatctgcagctgcctgtgtgcagacatgtgatttgaaagagttcacacct 780
Db 1919 CAGTGGAAAGATTTGATTCAGTTCGCTGTGCGAACAATGTATTTGGAAGAGTTTCAGACT 1978
Qy 781 gctgtgtgcgcactgcagaaagcaltgcgttgcctgtgtgcctgtgtgcacaccttgcctga 840
Db 1979 GTCTGTGTGCGCACTGCACAGCATGCGTTGCTGTGCTGCTGTGCACACCTCTGCTGGA 2038
Qy 841 aagtgctcattccgggagacacatgccttcgagagctctgtcggatgggagaaatg 900
Db 2039 AAGTGTCTATTCCTGGGGACACCATGCTCTGCGAGGCTCTGTGCTGGATGGGAAAGATG 2098
Qy 901 ccacctcctctgatacagaagccaccttgaagaatgtgttgaaagagaaagcagtgagaa 960
Db 2099 CCACCTCTCTGATACATGAAAGCCACCTGGAAGATGTTGGAAGAGGAAGCATGTGAAA 2158
Qy 961 agacacacagcacaacgctcccaagccatcagctgtggagatg 1001
Db 2159 AGACACACAGCACAACGCTCCCAAGCATCAGCTGCGGATG 2199

RESULT 3
BC004158 3006 bp mRNA PRI 16-MAR-2001
LOCUS Homo sapiens, putative prostate cancer susceptibility protein,
DEFINITION clone MGC:2441, mRNA, complete cds.
ACCESSION BC004158
VERSION BC004158.1 GI:13278770
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3006)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCM/DMP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLUM)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan
Clone distribution: MGC clone distribution information can be found
```

```
through the I.M.A.G.E. Consortium/ILLUM at: http://image.llnl.gov
Series: IPAL Plate: 2 Row: c Column: 13.
FEATURES
source
1. 3006
/organism="Homo sapiens"
/db_xref="locusid:60528"
/db_xref="taxon:9606"
/clone="MGC:2441 IMAGE:2820640"
/tissue_type="Lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"
/lab_host="DH10B-R"
/notice="Vector: pOTB7"
70..2550
/codon_start=1
/product="putative prostate cancer susceptibility protein"
/protein_id="AAH04158.1"
/db_xref="GI:13278771"
/translation="MMAICSLRNSAGRTMSOGRTTISQAPARRRPRKDPRLRIHRTRE
KRSGSGSGGPNIVLYOVVAGSRDSCALVSEFNNRYLPNCGEVQRLMOEHKLV
ARLDNIPLTRHMSNVGSLGMLILTKETGKCVLSPDPLEKYLAIKIFSGPLKG
IELAVRPHSAPEYEDETMYQIPLHSORRGKHQPMOSPRRPLSRSPERSSEIN
ENEPHLPHGVSORRGVDRSSLYVAFICKLHKRGNFLYIAKKEKGLPGVTAIAPITA
AVKDGKSTTHGREGREILAEELCTPPDPGAAFPVVECPDESFTOPICENATPRYGRKD
APVALVYHMAFASVLYDSRYQMMERREPDQHLVLNENCAVHNLSHKIOTQNLIL
HPDIPEPLTFSRCKKEGPTLSVPVQSECLIKYLRPRRQDAIITCNPEEIVEA
LQLPFOOSVOEYRRAADGPAPAEKRSQYEIIFLGTGSAIIPKIRNVASTLVNISP
DTSLDDEGTFGOLCRHYGDQVDRVLTAAVAVSHLADHHTGSLILOERL
ASICKPLPLILVVA.PNOLKAMLOOYHNOGCVLHISMIPAKIGOLEGSEISSPVERL
ISSILRTDLEFPOTCLVRHCKNAFGCALVTSMTKYYISDTPRCALVYMGMDATL
LIHRAVTEDEGEAEVETHTSTQSALSVGRMAEFTMLNHFQRYAKVLPSPNS
EKGVAVRDMKVCFGDEPTMKLIPPLKALFAGDIIEMEERERRELIQVRAALLSRE
LAGLIEGDEPQOKRAHTEEPQAKVRAQ"
BASE COUNT 742 a 808 c 858 g 598 t
ORIGIN

Query Match 98.6%; Score 987.4; DB 91; Length 3006;
Best Local Similarity 99.8%; Pred. No. 2.8e-261;
Matches 999; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
```



```

Db 1689 TCGTGTCTTGTGTCCACCT-GGACGACATGACACACAGGGGTTCCAGATATCTTC 1747
QY 541 tgcagagagaagcgccttgatcttggagaagccgtccaccttgctggtgtg 600
Db 1748 TGCAGAGAGAACCGCTTGGCATCTTTGGGAAAGCGCTTCAACCCCTTCTGCTGTG 1807
QY 601 ccccaacacgcctcaagccttgctcagagatcccaacacagtgccaagagctctgc 660
Db 1808 CCCCCAACCGCTCAAGGCTGGCTCCAGAGTACCAACACAGTCCAGAGGCTCTGC 1867
QY 661 accacatcagatgatctcctgcacaatgacctcaagagaagggctgagatccagtcgt 720
Db 1868 ACACATCATGATGATTCCTGCCAATGCTTCAGAGAGGGGCTGAGATCTCCAGTCTG 1927
QY 721 cagtgagaagaattgatcagctgctgtgcaacatgtgatttggagaagattccaagcct 780
Db 1928 CAGTGGAAAGATGATGATTCCTGCTTGGCAACATGTGATTGGAAGAGTTTCAGACT 1987
QY 781 gctcgttgaggcactcaagcctgttgctgtgctgctgctgctgctgctgctgctgga 840
Db 1988 GTCTGTGGCGGCACTGCACATGCGCTTGGCTGTGCGCTGTGCACACCTCTGCTGGA 2047
QY 841 aagtgcttctatccggggaacacatgcctcgcagagctcgtcccgagatggggaagatg 900
Db 2048 AAGTGTCTATTCGGGGGACACCATGCTCCGACAGCTCTGCTCGGATGGGAAAGATG 2107
QY 901 ccaacctctgtatcatgaaagccacctgtgaagatggtttgagaagagaagcagttgaaa 960
Db 2108 CCACCTCTCGATATCATGAAAGCCACCTGTGAAGATGTTGGAAGAGAGAGAGTGGAAA 2167
QY 961 agacacacagcacaagctcccaagcctcagctggtggagatg 1001
Db 2168 AGACACACAGCACAACGCTCCCAAGCCTACGCTGGGGGATG 2208

RESULT 4
AF308698
LOCUS AF308698 2908 bp mRNA PRI 27-FEB-2001
DEFINITION Pan troglodytes ELAC2 mRNA, complete cds.
ACCESSION AF308698
VERSION AF308698.1 GI:10946496
KEYWORDS
SOURCE
ORGANISM
Chimpanzee
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
1 (bases 1 to 2908)
Tavtigian,S.V., Sismard,J., Teng,D.H.F., Abitia,V., Baumgard,M.,
Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
Destrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
Ghaflari,S., Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N.,
Laitly,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
Neuhansen,S., Kommens,J. and Cannon-Albright,L.A.
A candidate prostate cancer susceptibility gene at chromosome 17p
Nat. Genet. 27 (2), 172-180 (2001)
11175785
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TAVITIGIAN,S.V., SISMARD,J., TENG,D.H.F., BAUMGARD,M., DAYANANTH,P.,
GHAFLARI,S., GUPTA,J.S., HU,R., ILIEV,D., JANACKI,T., KORT,E.N.,
LAITY,K., LEAVITT,A., LEBLANC,G., MCARTHUR-MORRISON,J.,
PEDERSON,A., PENN,B., PETERSON,K.T., REID,J.E., RICHARDS,S.,
SCHROEDER,M., SMITH,R., SNYDER,S.C., SWEDLUND,B., SWENSEN,J.,
THOMAS,A., TRANCHANT,M., WOODLAND,A.M., LABRIE,F., SKOLNICK,M.H.,
NEUHANSEN,S., KOMMENS,J. and CANNON-ALBRIGHT,L.A.
Pan troglodytes ortholog of human HPC2/ELAC2
Unpublished
3 (bases 1 to 2908)
TAVITIGIAN,S.V., SISMARD,J., TENG,D.H.F., BAUMGARD,M., DAYANANTH,P.,

```

```

Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
Ghaflari,S., Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N.,
Laitly,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
Neuhansen,S., Kommens,J. and Cannon-Albright,L.A.
Direct Submission
Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
Lake City, UT 84103, USA
FEATURES
source
location/Qualifiers
1..2908
/organism="Pan troglodytes"
/db_xref="taxon:9598"
1..2481
/codon_start=1
/product="ELAC2"
/protein_id="AAG24920.1"
/db_xref="GI:10946497"
/translation="MVALCSILRSAGRTMSQRTISOAPRRERPRKDPRLRTRE
KRPSGSGSPNTVYLQVVAAGSRDGAALVYSENRILFNCBGIORLMDBRLKY
ARLDNIFLTRMHMSNVGSLGMLTLTKELGIPCVLSGPOLKYLEALIKIFSGPLK
IELAVRPHSAPXEDEFTMTVYQIPHSEDRGRKHQWQSPERPLSRLESBSDESSE
ENEPHLPHGVSORGVSDSLVYAFICKLHLKGNFLVLAKEKMGAPVGAATAPIIA
AVMDKSTHHEGRELIAELCTPPDGAFAVVECPDESFIOICENATFOROGKAD
APVALYVHNAPESVLYDSRYQOMMERFQDTQILYINENCASHNIRSHKIQTQMLI
HDIFPLTFSFCKRKGPPLSVPMVQGECLIKQLPRRMDALITONPEFTILEA
LQIPNPOQSVFERRSADODPAPEAKRSQYPELIIFGTGSAIPMKIRNATLVNISP
DTSLIDCGEFTGQLCRHYGDVDVRLTLAAVEFSHLHADHTGLINLLOREARL
ASLGRPHLLVAVAPNOLKAMLODYNOOCEVHNHISMIIPAKTLOGEAEFISVAPRL
ISLRLTCDLEFQRTLVNRCKAFCAALVHNSGKVVYSGDMPCEALVRMKDQTL
LHETATLEDGLEBEAVEKHTSTTSQATISGMRNNAEFTMLNHSQATVAPYLPSPFN
EKVGAFFDMKVCEDFATMPKLIPLKALFADIDEMERREKRLROYRAALLSRE
LAGLEDGEPPOOKRAHTEBPQAKVRAQ"
BASE COUNT 712 a 788 c 819 g 589 t
ORIGIN
Query Match 97.2%; Score 973; DB 89; Length 2908;
Best Local Similarity 98.9%; Pred. No. 2, 6e-257;
Matches 990; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 1 tgcgagcacaagaattcaaaacccagctcaacatcaccacgggaacttccccgct 60
Db 1140 TCGAGCGCACAAAGATTTAAACCCAGCTCAACCTCATCCACCGGACATCTTCCCCTGCT 1199
QY 61 caccagttccgcctgtaagaagaagagcccaacctcagtggtgccattcagagtga 120
Db 1200 CACAGTTTCCCTGTAAAGAAAGAGGGCCCAACCTCAGTGTGCCATGTGACGGGTGA 1259
QY 121 atgcctcctaagtaacagctcgcgtcccaagagaggtgcaagaaggatccattatc 180
Db 1260 ATGCTTCCTCAAGTACACAGCTCCGCTCCAGAGAGGAGTGGCAGAGGATGCCATTATTAC 1319
QY 181 ttgcaatcccggaagatcctagtgagcgctgagcttcccaacttccaagaagagct 240
Db 1320 TTGCAATCCGAGGAATTCATTAATGTAGCGCTGCACCTTCCCAACTTCAGCAGAGTGT 1379
QY 241 gcaggaatagaagaagtgctgcagagagcccaagcccgcaagagaagaagaagtcagta 300
Db 1380 GCAAGATACAGAGAGAGTGCGCAGAGCGGCCACGCCACAGAGAAAAGAACTCAGTA 1439
QY 301 cccgaatacatcttcttggaaacaggtctgcacatcccgatgaagaatcgaaatgtcag 360
Db 1440 CCCAGAATCATCTTCTTGGAAACAGCGTGCATCCCATGAAGATTCGAATGTCAAG 1499
QY 361 tgcacacgtgttaactaagaagcccgacagctctctcttgacgtgtgtgaagcac 420
Db 1500 TGCACACTTGTCAACATTAAGCCCGACAGCTCTCTCTACTGTGACTGTGTGAGGCGAC 1559
QY 421 attggacagctgtccgcatattacagagacaggtgtaagaaggctccggcacccctgac 480
Db 1560 GTTTGGGACACTGTGCCGTATTACGAGACACAGTGTGACAGGGTCTGGGGACCCCTGGC 1619

```

OY	481	tgcctgtgcttgctgccaccctggcaagcagatcacccaaggcgcttgcgaagtattcgc	540
Db	1620	TGCTGTGTTGGTGCACACT-GCACGGCAGATCACACCACGGCGTTGCTAAATATCTTCG	1678
OY	541	tgcagaagaacagcgcccttgagcatctttggaaaagcgcttaaccttgtcgtgtgtg	600
Db	1679	TGCAAGAGAACGACGACTTGGAATCTTTTGGAAAAGCCITTCACCTTTGTGTGGTTG	1738
OY	601	cccccaaccagctcaaaagcttgcttcagcagtatcacacaacagttccagaggtctcgc	660
Db	1739	CCCCAACACGCTCAAAGCGCTGGCTGCACAGTACCAACAACAGTGCCAGGAGTCTCTGC	1798
OY	661	accacatcagtatgatctctgcaccaaatccttcagaagaaggcgtagatctccagtcctg	720
Db	1799	ACCACATCATGTATGTATTCTCTGCCAATATGCTTCAGGAAGGGCGTAGATCTCCAGTCC	1858
OY	721	cagttgaagaagtatgatcagttcgtcgtgtgcgaacatgtatattggaagaagttcaact	780
Db	1859	CAGTGAAGAAGATTGATCACTTGGCTGTTCCGAACATGTATTTGGAAAGTTTCAGACCT	1918
OY	781	gtctgtgtcgcgcacgtcycaagcatgcgtttgtcgtgtgtgcgtgtgcaacctctygtga	840
Db	1919	GCTGTGTGGCGCATGTCAAGCATGCGTTGTGGCTGTGCCTGGTGACACCTCTGGCTGGA	1978
OY	841	aagtgctcattccgcgggagcacacatgcctctgagaagctctgtgtccgatagggaagatg	900
Db	1979	AAGTGTCTATTCCGGGACACCATCTCCCTGCGAAGCCTCTGTCTCCGATGGGGAAGATG	2038
OY	901	ccacctctctgatcatgtgacgcaccttgaaagatggttttggaaggaagcagttgaa	960
Db	2039	CCACCTCTCTGATACATGTAAGCCACCTCTGGAAGACGCTTTGGAAGAAGAACGATGGAA	2098
OY	961	agacacacagcacaacgltccaaagccatcagtcgtggagtg	1001
Db	2099	AGACACACAGCAACGTCCTCCAAGCCATCAGCGTGGGATG	2139
RESULT 5			
AK001392			
LOCUS	AK001392	2976 bp mRNA	PRI 22-FEB-2000
DEFINITION	Homo sapiens CDNA FLJ10530 fls, clone NT2RP2000985.		
ACCESSION	AK001392		
VERSION	AK001392.1 GI:7022621		
KEYWORDS	oligo capping; fls (full insert sequence).		
SOURCE	Homo sapiens testiculocarcinoma cell line:NT2 CDNA to mRNA, clone_lib:NT2RP2 Clone:NT2RP2000985.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (sites)		
AUTHORS	Isogai,T., Oca,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiraori,A., Sudo,H., Wagatsuma,M., Hosokiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiuchi,S., Watanabe,S., Kimura,K., Muraoka,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu A., Nakamura,Y., Negishi,K., Masuno,Y., Nimomiya,K. and Iwayanagi,T.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	NEDO human cDNA sequencing project unpublished (2000)		
REFERENCE	2 (bases 1 to 2976) Isogai,T. and Otsuki,T.		
AUTHORS	Isogai,T. and Otsuki,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-FEB-2000) to the DBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@rhl.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952) NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: CDNA full insert sequencing: Research Association for Biotechnology: CDNA library construction, 5'- & 3'-end one pass sequencing and clone selection Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science,		

FEATURES	University of Tokyo.
source	Location/Qualifiers
	1. .2976
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/cell_line="NT2"
	/cell_type="teratocarcinoma"
	/clone_id="NT2RP200985"
	/note="cloning vector: PME18SF ₃ ; mRNA from NT2 neuronal precursor cells after 2 weeks retinoic acid (RA) induction."
	75. .2639
	/note="unnamed protein product"
	/cdon_start=1
	/protein_id="PAA91666.1"
	/db_xref="gi:7022622"
	/translation="MMALSLIRSAAGRTMSOGRTTISOAPRRRPRKDDLRHLRREE KRGSQSGSGGNPTYYILOVVAAGSDSALATVPSENPRLFPNCEGVORLMOEKRLV ARLUNIFLTRHMSNVGSLGSMILITLTELGPVKVLSRPPOLELTLDAITISGPLG IELAMRPHSAPEYEDMTVYOIPIHSEORGRKHQPMQSPERLSRLSPERSDSESN ENELHPHVSORGRVDSILVAFIEIKLIRKNFVLIAKLEMGILVGAALIPITA AVKRGKSTHGREILAEELCTPPDPAAFVYVEPDESFLOPCENATFORQGRAD APVALYHMAASVINDSRXOQMERGPRDQHYIENCAASVHNLRSHTIQNLIL HPDLPILSTRCKEGPTLSVPMVOGELKLTQLRPREMRDAITTCNBERFVEAL QLPNNQSVDEYRRSADDPAPAPAKRSQYPIITFLGSGASIPKIRNVAATLVNISPD TSLDDQSEGTGOLCRHYGDQVDRVLTAAVSHLMDHHTGSLISLORRALA SLGRLPHLIVANPOLKHAQYINOCQELIHSIIPAKLOEGAEISPAVERLI SSLRTGDLDEFORCLVRHCKHAGCALVHTSGMKVYVSGDTPCEALVVRGDAITLL IHEATLEDLGEYEDVEKTHSTTSOATISVGMNMAEFTIMNHFSQVAKVPLSPNSEE KVGAFTLMEKGLDEDFTPMKLIPLPALRAGDI EEMDERREKELQVAAALSRLL AGGLEDSLKSGTQPSHRPRSEPEDLGDELRLVCVCPTHAPVSLALGRS"
BASE COUNT	712 a 807 c 856 g 601 t
ORIGIN	
Query Match	97.0%; Score 971.4; DB 89; Length 2976;
Best Local Similarity	99.5%; Pred. No. 7.1e-257;
Matches 996; Conservative	0; Mismatches 1; Indels 4; Gaps 2;
1	tgcagagcaccaagattcaaacccagctcaactatcaccccgagcatctccctgct 60
1214	tgcgagccacaaagattcaaacccagctcaactatcaccccgagcatctccctgct 1273
61	caccagattccgcttaagaagaagagcccccaccctcagtggtgccatggttcaaggtga 120
1274	caccagattccgcttaagaagaagagagcccccaccctcagtggtgccatggttcaaggtga 1330
121	atgcctctctcaagtaaccagctcgtctcccaaggaaggagtggcagaaggatgccaattatcac 180
1331	atgcctctctcaagtaaccagctcgtctcccaaggaaggagtggcagaaggatgccaattatcac 1390
181	ttgcaatctctgaggaatcatcagttgaagcgctgtaagcttcccaacttcagccagaaggt 240
1391	ttgcaatctctgaggaatcatcagttgaagcgctgtaagcttcccaacttcagccagaaggt 1450
241	gcagaagtaacaagaagatgtagcagaagcgcccaagccccaagaagaagaagaagtcagta 300
1451	gcagaagtaacaagaagatgtagcagaagcgcccaagccccaagaagaagaagaagtcagta 1510
301	cccaagaatcatctctctcttggaacagggctgtgcatacccgatgaagattcgaaatgtcag 360
1511	cccaagaatcatctctctcttggaacagggctgtgcatacccgatgaagattcgaaatgtcag 1570
361	tgcacacttgttcaacaataagccccgaagaagcttctgtctactcgtgaactgtgtgagggac 420
1571	tgcacacttgttcaacaataagccccgaagaagcttctgtctactcgtgaactgtgtgagggac 1630
421	atttgagcagctgtgcgttcaattacggagacaaggttgacaagggctgcttgggacacccctggc 480
1631	atttgagcagctgtgcgttcaattacggagacaaggttgacaagggctgcttgggacacccctggc 1690
481	tgcctgtgtctgtgtcccaacttgacgcagaagatcaaccaacagggcttgccaagtatcttgc 540

DB	1691	TCGTCGTGTTGGTCCCACT	GCACGAGATACACACGCGGCTTGCCAACTATCTTGC	1749
QY	541	tcgacagagaacgcgcctctggcatctcttgggaagaagcgccttcacccctcttgcgtgattg		600
Db	1750	TGCACAGAGAACGCGCCTTGGGATCTTTGGGAAACCGCTTACCCCTTTCGCGTGATTG		1809
QY	601	cccccaaccagctcaaaagccttgctcagcagctacccaacaacagtgccaagaagctctgc		660
Db	1810	CCCCCAACCACTCTAAAGCCCTGGCTCCAGAGTACTACAAACAGTGCAGAGAGGCTCTGC		1869
QY	661	accacatcagatgatctctctccaaatgccttcagaagaaggagcttaagctctccactctg		720
Db	1870	ACCACATCAGATGATGTTCTCTGCCAAATGCTTTCAGAAAGGGGCTAGATCTCCACTCTCG		1929
QY	721	cagtgaaaagatttgatcagttcagttctgttcgcagacatgtgatttggaagaagttcagact		780
Db	1930	CAGTGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT		1989
QY	781	gtctggtgcgcgcacgtcgaagcagtcgcttgctgtgctgtgctgtgctgtgctgtgctgtg		840
Db	1990	GTCCTGTCGCGGACACGACGATGCGGTTGGCTGTGCGCTGTGTCGACACCTGTGACTGGA		2049
QY	841	aagtgatctatctccgggagaccatcagccctgtggaagctctgtgctccgattgggaagatg		900
Db	2050	AAGTGCTATTTCCGGGACACACATGCGCTGGAGGCTCTGCTCGGATGGGGAAGATG		2109
QY	901	ccaccctctgatacatgaaagccaccctgggaagatgatttggaagaaggaaagcagtgaaa		960
Db	2110	CCACCTCTCTGATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT		2169
QY	961	agacacacagacaacacgctcccaagcattcagcgttggagatg		1001
Db	2170	AGACACACAGACAACACGCTCCCAAGCCATCAGCGTGGGATG		2210
RESULT	6			
LOCUS	AF308694	2893 bp	mRNA	PRI
DEFINITION	Gorilla gorilla ELAC2 (ELAC2) mRNA, complete cds.			27-FEB-2001
ACCESSION	AF308694			
VERSION	AF308694.1	GI:10946488		
KEYWORDS				
SOURCE	gorilla.			
ORGANISM	Gorilla gorilla			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Gorilla.			
AUTHORS	1 (bases 1 to 2893) Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M., Dumont,M., Farham,J.M., Frank,D., Frye,C., Ghaffari,S., Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J., Peterson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhansen,S., Rommens,J. and Cannon-Albright,L.A.			
TITLE	A candidate prostate cancer susceptibility gene at chromosome 17p			
JOURNAL	Nat. Genet. 27 (2), 172-180 (2001)			
REFERENCE	11175785			
AUTHORS	2 (bases 1 to 2893) Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M., Dumont,M., Farham,J.M., Frank,D., Frye,C., Ghaffari,S., Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J., Peterson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhansen,S., Rommens,J. and Cannon-Albright,L.A.			
TITLE	Gorilla gorilla ortholog of human HPC2/ELAC2			
JOURNAL	unpublished			
REFERENCE	3 (bases 1 to 2893)			

FEATURES	source	Location/Qualifiers
gene	1. .2893	/organism="Gorilla gorilla"
	1. .2893	/db_xref="taxon:9593"
CDS	1. .2481	/gene="ELAC2"
	1. .2481	/note="similar to Homo sapiens HPC2/ELAC2"
	1	/codon_start=1
	1	/product="ELAC2"
	1	/protein_id="AAC294916.1"
	1	/db_xref="GI:10946483"
	1	/translation="MMAISCLSLRSGAGRTMSQGTITISQAPARPRPKDPLRLHRTREKRGSCGSGGPNVYQLQVVAAGSDSQAALVSEFNPRTYLFNCEGVQALMOEHLKVVLRLNIFLRTMHMNSVGLSGMLITLGTGKPKQVPLEYATIKFIKFSPLKGIELAVRPASPEYEDMTMYQIPIHSEORGRKQVMSPPRLSRLSPRSSSENENEPHLPHGVSQRRGVDDSSLYVAFCTKLHLKRNPLYLAKEMGLVGTATAPITAAVKQKGSITHEGREILAEELCTPPDPAAGPAAVVECPDESFIPICENATFORGRADAPVAVLVMAPEVALVDSRYQOMMERARFPDTHLVENCSVYNLSKSHIQDNLNLIHPDLPLTSPPEPKKEEPLTSPVAGCECLLYLRPREREMORDAITTCDEEPIYEAALQNPVQSVQEVRYODVDPAPAKERSQYDEIFLGTGSAIPKIKNSVATLVNISPDLSLLDGEESTFGOLCRHNGDQVDRILGTAAAFVSHLADHTGKLTILLQREOMLASELKRPLHPLVAPSQLKMANLOLYHMOCEVLSHIMTIPAKCLQEBALISPAVELLISLRLTQDEEFQTCVLRHCKHAFGALVATSWMKVYSGDTPMPCALVMSPAVELLISHRTLEDGLDEEAAVERKSTTSQALSVGRNMAEIMLNHFSQRAKVPLEPNPNLEKVVADHMKVCGDEPTPTMKLIPKALFAQDIEMEEERERERELROYRAALLSGEELAGLEGEPEQKRAHTEEPQAKKVRQ"
BASE COUNT	705 a	787 c 815 g 586 t
ORIGIN		
Query Match	96.4%	Score 965; DB 89; Length 2893;
Best Local Similarity	98.4%	Pred. No. 4,1e-255;
Matches	985; Conservative	0; Mismatches 15; Indels 1; Gaps 1;
1	1	1
11440	TCGCAGGCCAACGATTCAAAACCACCTCACTCACTCAACCCGCGACATCTTCCCTGCT	1199
61	caccagtttcgcgttaagaagaagggcccaacctcaagtgtgccatggttaagatga	120
1200	CACCGATTTCCTCCCTTAAGAAAGAGAGGCCCCACCTCAGTGTGCCAAGTTCAAGGTGA	1259
121	atgctctctcaagtaaccagctctccctcccaagaaggagcggcagaagagatgcatattac	180
1260	ATGCTCTCTCAAGTACCAAGCTCCGTCCTCCAGGAGGGAATGGCAGAGGATGCCATTATCAC	1319
181	ttgcaatcttgagaattcatatgattgtaggcgtgtgacgttcccaacttccagcagaagcgt	240
1320	TTGCAATCTTGAGGAATTGATTGTTGAGGCGGTGACGCTTCCCACTTCCAGACAGAGT	1379
241	gcaggagtaacaggaagatgcgacgaagcgcagcccccagcagagaagaagaagtcaagta	300
1380	GCAGGAGTACAGAGAGAGAGTGTGTCAGAGAGCTCCACCCACAGAGAGAAAGAGTACACTA	1439
301	cccagaatacatcttcctcttgagaagaggcttggccatcccgatgaagaattcgaaatgtcag	360
1440	CCCAGAAATCATCTTCTTGTGACACAGGGCTGTGCCATCCCATGAAGATTCCGAATGTGAG	1499
361	tgccacactgttcaacataaagcccgacacagctctctgtactgagacttggtaagggcac	420


```

QY 62 accatttccgcctgaagaagaaggccaccctcagctgtgcccattggttcagggtgaa 121
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1203 ACCAGCTTTATAGTAAGGAGGAGGATCCACCTTCAGCTGCACAGTTCCGGGGTAA 1202
QY 122 tgcctcccaagtlaccacccgtccgltcccaaggaggagtgagagaggaatgcattact 181
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1263 TGCCTCCTCAAGTATCAGCTCCGCCCAAGAGAGAGTGGAGAGGATACCAACCCGAC 1322
QY 182 tgcattccctgaagaatattcattgtagggcctgcagcttcccaattccagcagcgtg 241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1323 TGCATAATCTGATTAATTCAATGAGGCTTGAGAGCTCCCGATTCCAGAGAGAGTGTG 1382
QY 242 caggagtlacagagagagtgtagcagagcagcagccagccagcagagaaagaatcagtlac 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1383 GAGAGATATGGAGAGACGTGCGAGAAACCACCCACAGAGAGAGAGAGAGAGAGAGTAT 1442
QY 302 ccagaaatcatctctcttggaacaaggctctgcattcccgatgaagatcgaattcagtlact 361
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1443 CCGTAATTTGTTCTTCGGGTACGGGGTGTGCATCCCAATGAAGATCCGAATGTCTAGT 1502
QY 362 gccacactgtcaacataagcccgcagacgctctctgactgagctgtggtgaggacaa 421
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1503 TCCACACTCTCTCAACTTAACCTTGACAAAGTCAAGTCTCTCGATGTGTGAGAGAGCACT 1562
QY 422 ttggagcagctgtgcccattatcagagaccaggtgagacagaggttccctggcaccctgct 481
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1563 TTGGGGCAGTGTGGCCCTCTTTACGGACACCAATAGACCGAGCTTATGACAGCTTACG 1622
QY 482 gctgtgttctgtccacactgtgcagcagagatcaacacagggcttgcagatctgct 541
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1623 GCGTGTGTTTGTGCCACCT - GCACGCCGACACACACAGGGGGTGTGTAATATCTTGCT 1681
QY 542 gcagagagaagcgcgcttgagcctcttgaggaaagcgcgttccacccttgcgtgtgtgtgc 601
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1682 GCAGAGAGACACAGCGTGTGCATCTGTGGGAAACCCCTTCAGCCCTTGCTGTGGTGGCG 1741
QY 602 ccccaacacagctcaaaagcgcgtctccagcaggtaccacacagtcagagagagcctgca 661
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1742 TCCATACCACTCAAGGCTGTGGCTGCAGCAGTATCACACCACTGCCAGGAGATTCTGCA 1801
QY 662 ccacatcagatgattctcttcccaatgctctcaggaagggtctgagatctccagtlctgc 721
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1802 CCACGTCAGATGATTCCTCCCAATGCTTCAGAAAGGGGAGAGGCTCTCCCAATACTAC 1861
QY 722 agtggagaagaatgattcagtlcgtgtgttgagaaatgattggaagaagtlcagactgt 781
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1862 ATTGGAAAGGCTGATAGCTTGTCTGTGGAAACATGTGACTTGAAGAAATTTTCAGACCTG 1921
QY 782 tctggtggagcagctgcaagatgctgttgctgtggtgtggtgacacacctctgctgtgaa 841
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1922 CCGTGTACGCGACTGCAGAGATGCTTTGGCTGTGCACTGTGATCATTCCTGGCTGAA 1981
QY 842 agtgcctattcggggagacacacatgcctctgagagcctctgctcggatlgggaaagatgc 901
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1982 AGTCGTCTACTCGGGGATATACATGCCCTGTGAGGCTCTGTCAGATGGGAGAAAGATGC 2041
QY 902 caccctctctgatacatgaagcacccttgagaagatglttggaagaggaaagcagtlgaa 961
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2042 CACCTCTCTGATACATGAACCCACTCTGGAGATGCTTGAAGAGAGACAGTAGAGAA 2101
QY 962 gacacagagacaagctcccaagccatcagcgtgggagatg 1001
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2102 GACACACAGCACCACTCCAGGCCATTAAATGTGGGATG 2141

```

```

RESULT 8
LOCUS AC005277 118788 bp DNA PRI 23-JUL-1998
DEFINITION Homo sapiens chromosome 17, clone hRPK.597_M.12, complete sequence.
ACCESSION AC005277
VERSION AC005277.1 GI:3337311
KEYWORDS HTG.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 118788)
AUTHORS Birren,B., Fasnman,K., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.597_M.12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 118788)
AUTHORS Birren,B., Fasnman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckert,J., Benm,J., Boatn,C.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
Gensheimer,S., Geraghty,K., Gilmartin,T., Grant,G., Hagos,B.,
Harris,K., Horton,L., Howland,J.C., Hul,L., Jacotot,L., Kann,L.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Nechman,A., Nefl,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Strickland,C., Subramanian,A., Tortorella-Miller,I., Vassiliev,H.,
Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
Zody,M.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 118788)
AUTHORS Birren,B., Fasnman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckert,J., Benm,J., Boatn,C.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
Gensheimer,S., Geraghty,K., Gilmartin,T., Grant,G., Hagos,B.,
Harris,K., Horton,L., Howland,J.C., Hul,L., Jacotot,L., Kann,L.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Nechman,A., Nefl,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Strickland,C., Subramanian,A., Tortorella-Miller,I., Vassiliev,H.,
Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
Zody,M.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jul 23, 1998 this sequence version replaced gi:3335015.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Only the first 118.8 kilobases of this clone are being submitted.
The remainder overlaps accession number AC005274 (WICGR project
L350).
FEATURES
source
1. 118788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK.597_M.12"
/clone_id="RPCL-11 human BAC library"
/map="17"
/chromosome="17"
/region 795..884
/rpt_family="MIR"
/complement(932..1000)
/rpt_family="MIR"
2044..2076
/rpt_family="AT-rich"
2237..2713
/rpt_family="MER34"
2718..3477
/rpt_family="L2"
/complement(3478..3821)
/rpt_family="MLT1A1"

```

```
repeat_region      3822..4007
/rpl_family="L2"
complement(4015..4281)
repeat_region      /rpl_family="MER89"
complement(4282..4321)
repeat_region      /rpl_family="(TAA)n"
complement(4328..4608)
repeat_region      /rpl_family="Alusx"
complement(4609..4895)
repeat_region      /rpl_family="MER89"
complement(4896..5186)
repeat_region      /rpl_family="Alusg"
5219..5479
repeat_region      /rpl_family="L2"
complement(5481..5740)
repeat_region      /rpl_family="MER34"
complement(5963..6275)
repeat_region      /rpl_family="Alusg"
complement(6411..6494)
repeat_region      /rpl_family="MER5A"
6830..7013
repeat_region      /rpl_family="MLT1C"
7014..7318
repeat_region      /rpl_family="AluDb"
7319..7581
repeat_region      /rpl_family="MLT1C"
7721..8032
repeat_region      /rpl_family="AluDb"
8039..8333
repeat_region      /rpl_family="Alusg"
9431..9460
repeat_region      /rpl_family="AT-rich"
9662..9829
repeat_region      /rpl_family="MIR"
complement(11302..11728)
repeat_region      /rpl_family="MER39"
complement(11824..11884)
repeat_region      /rpl_family="LTR16C"
complement(11928..12108)
repeat_region      /rpl_family="LTR16C"
12453..12725
repeat_region      /rpl_family="AluDb"
13377..13785
repeat_region      /rpl_family="MSTB"
14118..14153
repeat_region      /rpl_family="(CA)n"
complement(15320..15620)
repeat_region      /rpl_family="AluDb"
15683..15759
repeat_region      /rpl_family="(CA)n"
complement(15838..15893)
repeat_region      /rpl_family="MER91A"
complement(15915..16080)
repeat_region      /rpl_family="MER91A"
complement(16474..16544)
repeat_region      /rpl_family="L2"
17907..17920
repeat_region      /rpl_family="L1ME"
18271..18311
repeat_region      /rpl_family="L2"
complement(18368..18529)
repeat_region      /rpl_family="AluDb"
complement(18530..18560)
repeat_region      /rpl_family="(CA)n"
complement(18561..18681)
repeat_region      /rpl_family="AluDb"
18760..19065
repeat_region      /rpl_family="Alusg"
complement(19638..19896)
repeat_region      /rpl_family="MIR"
complement(19971..20260)
repeat_region      /rpl_family="L1ME3A"
20263..20552
```

```
repeat_region      /rpl_family="MER58B"
complement(20553..20816)
repeat_region      /rpl_family="L1ME3A"
complement(20821..21069)
repeat_region      /rpl_family="L1M4"
complement(21071..21100)
repeat_region      /rpl_family="(TAA)n"
complement(21101..21338)
repeat_region      /rpl_family="Alusx"
complement(21348..21598)
repeat_region      /rpl_family="L1M4"
21799..22104
repeat_region      /rpl_family="AluDb"
22151..22562
repeat_region      /rpl_family="MSTA"
23978..24287
repeat_region      /rpl_family="Alusg"
24852..25199
repeat_region      /rpl_family="L2"
complement(25697..25985)
repeat_region      /rpl_family="MLT1D"
25816..25895
unsure
/notice="Single-stranded coverage."
repeat_region      complement(25994..26099)
repeat_region      /rpl_family="purine-rich"
complement(26122..26291)
repeat_region      /rpl_family="MLT1D"
26312..26337
repeat_region      /rpl_family="(CAAA)n"
26456..26480
repeat_region      /rpl_family="(GGGA)n"
complement(26576..26781)
repeat_region      /rpl_family="MER5B"
complement(26912..26990)
repeat_region      /rpl_family="MER21B"
26991..27290
repeat_region      /rpl_family="Alusx"
complement(27291..27975)
repeat_region      /rpl_family="MER21B"
28034..28197
repeat_region      /rpl_family="LTR28"
28266..28795
repeat_region      /rpl_family="LTR28"
28843..28916
repeat_region      /rpl_family="LTR28"
28920..29005
repeat_region      /rpl_family="MIR"
29140..29233
```

Query Match 13.7%; Score 137; DB 85; Length 118788;
Best Local Similarity 74.3%; Pred. No. 5.3e-27;
Matches 225; Conservative 0; Mismatches 0; Indels 78; Gaps 1;

```
QY 668 cagatgatctcgtccaaatgccttcaggaaggagctgagatccagtcctgcagtgga 727
|||||
DB 101444 GAGTATGATTCCTCGCAAAATCGCCTTACAGGAAGGGCTGAGATCTCAGCTGCAGTGA 101503
|||||
QY 728 aagatgatcagttcgtcgttgccgaacatgatttgaa 768
|||||
DB 101504 AAGATTGATCAAGTTCGTGCGAACAATGATTTGGAAGGTAAGGGGCACAGCCGC 101563
|||||
QY 769 -----a 769
DB 101564 AGGCATCATGGGGGCGAGGTGGGGAGCAGAGCTGCAGAGCCCTCCAGCCCACTTTCA 101623
|||||
QY 770 gttcaagacctgctcgtgcyggaactgcaagacatgcatgcttgctgcytgatgacac 829
|||||
DB 101624 GTTTCAGACCTGCTCGTGGCGGACACTGCACATGCTGTTGGCTGTGCGTGTGCACAC 101683
|||||
QY 830 cctcgtcggaagagtgatctatctccgggagacacatgctcctgcgaggtctcgttcgcat 889
|||||
DB 101684 CTTCTGGCTGGAAGTGTCTATTCTCGGGGACACCATGCCCTGCGAGGCTGTGTCGGAT 101743
```


CDS
join(8470..8690,9077..9127,9266..9336,9948..10012,
10957..11014,12096..12164,12933..13064,14902..14960,
16609..16667,16746..16818,17370..17482,18671..18766,
21539..21677,21757..21842,23553..23671,24619..24715,
27291..27429,27763..27801,27930..28039,28553..28652,
28733..28853,29101..29179,29710..29854,30406..30624),
/gene="Elac2"
/note="24 exon form"
/codon_start=1
/product="putative prostate cancer susceptibility protein"
/protein_id="AAK29421.1"
/db_xref="GI:13540343"
/translation="MMALRSLLRPLGLRTMSQSGARRPPKDPRLHRTREKRGPG
GGPNVYLQVVAAGGRDAGALYFSEYNRLFNCGEGVORIMOBHKLKVARLDITFL
TRHMSNVGGLGMLTLKETGLPKCVLSGPPOLEYLAIRISGPALELAVRPH
SAPKEDTMTVYQVPIHSEKRCGKQSPRNSPMLSPROSSDSASAEKGPPE
DSAGANKRANGRPSLYAVFCKLHKGNFVLYKAEGLIPGNTAIAPIIAVAVD
CKSTYEGREIAEBELCTPDPGLVFIVCEPDGFIPLICENDFKRYQAEADAVK
LVNHAPESVLIHSRYQOMERFEPDTQHLILNENCPVHNRSHKIDOTSLHPDI
FPOLTSFKSEGSTLSVPTVREGCLVLPKREKQMDITLDCNTEITAELELP
SFOESVEYRKNOENPAERKRSQYPEIVETGSAIPMKIRNVSSTLVNISPDSY
LLDGEFTGOLCRHYGOQIDRVLCSTLAVVSHHAHHGILNILLORHAIASLG
KRPOLYVAPROLRAMYOYHNCOELIHHVSMTPACLOKGAIVSNTTERLSIL
LETCDLEFQCLVNRHCKHAFGCALVHSSKVVYSGDTMPCALVQMKDATTILHE
ATLEDGEEFAVEKHTSHSTSOAINVGMMAEFILNHSQYAKIPLFSPDNKVG
IAPDHMKVCFDPTVPKLIPLKALFAGDIEEWEVEREKRELRVRAALLTQADSP
EDREPOKRAHTDEHSPQSKKRAQ"
10957..11014,12096..12164,12933..13064,14902..14960,
16609..16667,16746..16818,17370..17482,18671..18766,
21539..21677,21757..21842,23553..23671,24619..24715,
27291..27429,27763..27801,27930..28039,28553..28652,
28733..28853,29101..29179,29710..29854,30406..30624,
31242..31277)
/gene="Elac2"
/note="25 exon form"
/codon_start=1
/product="putative prostate cancer susceptibility protein"
/protein_id="AAK29420.1"
/db_xref="GI:13540342"
/translation="MMALRSLLRPLGLRTMSQSGARRPPKDPRLHRTREKRGPG
GGPNVYLQVVAAGGRDAGALYFSEYNRLFNCGEGVORIMOBHKLKVARLDITFL
TRHMSNVGGLGMLTLKETGLPKCVLSGPPOLEYLAIRISGPALELAVRPH
SAPKEDTMTVYQVPIHSEKRCGKQSPRNSPMLSPROSSDSASAEKGPPE
DSAGANKRANGRPSLYAVFCKLHKGNFVLYKAEGLIPGNTAIAPIIAVAVD
CKSTYEGREIAEBELCTPDPGLVFIVCEPDGFIPLICENDFKRYQAEADAVK
LVNHAPESVLIHSRYQOMERFEPDTQHLILNENCPVHNRSHKIDOTSLHPDI
FPOLTSFKSEGSTLSVPTVREGCLVLPKREKQMDITLDCNTEITAELELP
SFOESVEYRKNOENPAERKRSQYPEIVETGSAIPMKIRNVSSTLVNISPDSY
LLDGEFTGOLCRHYGOQIDRVLCSTLAVVSHHAHHGILNILLORHAIASLG
KRPOLYVAPROLRAMYOYHNCOELIHHVSMTPACLOKGAIVSNTTERLSIL
LETCDLEFQCLVNRHCKHAFGCALVHSSKVVYSGDTMPCALVQMKDATTILHE
ATLEDGEEFAVEKHTSHSTSOAINVGMMAEFILNHSQYAKIPLFSPDNKVG
IAPDHMKVCFDPTVPKLIPLKALFAGDIEEWEVEREKRELRVRAALLTQADSP
EDREPOKRAHTDEHSPQSKKRAQ"

BASE COUNT 8641 a 8030 c 8142 g 9780 t
ORIGIN

Query Match 10.3%; Score 103; DB 94; Length 34593;
Best Local Similarity 88.2%; Pred. No. 1.4e-17;
Matches 112; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 769 agttcagaccgtctgtgtgagcactgaacatgcgttgcgtgtgcgtgtgcacaa 828
Db 28731 agtttcagaccgtctgtgtgagcactgaacatgcgttgcgtgtgcgtgtgcacaa 28730
Qy 829 cctctgctgtgaaagtgtctatctccgaggaacacatgccctgcgagcgtctgcgtccgga 888
Db 28791 cacttgctgtgaaagtgtctatctccgaggaacacatgccctgcgtgtgcgtgtgcacaa 28850
Qy 889 tggggaaa 895
|||||

Db 28851 TGGGTAA 28857
RESULT 11
SPAC1D4
LOCUS 31201 bp DNA PLN 15-OCT-1999
DEFINITION S.pombe chromosome I cosmid c1D4.
ACCESSION 269239
VERSION 269239.1 GI:1177333
KEYWORDS byrl; cct2; CDP-alcohol phosphatidyltransferase; chaparonin;
conjugation and sporulation; csk1; cyclin suppressing protein
kinase; DNA repair helicase; golgi peripheral membrane protein;
protein kinase; rad15; rhp3; serine threonine protein kinase; stel;
tcp1 beta subunit homologue; ttf2 LTR.
fission yeast.
SOURCE
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
1 (bases 1 to 31201)
Lye,G., Churcher,C.M., Barrell,B.G., Rajandream,M.A. and Walsby,S.V.
Direct Submission
Submitted (01-FEB-1995) Schizosaccharomyces pombe chromosome I
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CB10 1HQ. E-mail: barrell@sanger.ac.uk
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S-pombe>)
protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in Pombase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites.
CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
(complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid c1D4 is overlapped at the 5' end by cosmid c1f3 and at the
3' end by cosmid c22f3.
Location/Qualifiers
1..31201
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="I"
/clone="cosmid c1D4"
/map="IL"

misc_feature
gene
CDS
1..133
/note="The sequence of this region is derived from cosmid
c1f3, the true end of c1D4 is at position 134, while the
true end of c1f3 is at position 4137 in this sequence"
7..864
/gene="SPAC1F3.11"
/note="SPAC1D4.01"
7..864
/gene="SPAC1F3.11"
/note="SPAC1D4.01, len:285"
/codon_start=1
/product="hypothetical 33.0 kd protein"
/protein_id="CAB56125.1"
/db_xref="GI:5912524"
/translation="MRQIFGCTGTYMNGCYRQEGLELEPTTKSTRMNSIHKKSNSR
FRRRVGNKEKEDLELDNDIRLQALETKRRKIRNSIIIGINAEKLIQOETKKEK
QLNTPANEPHEANDTSAOSSKILIEAOLPTVEDREPAKQINEVDINTHLINFEVKLKOE


```

gene
complement(join(1088. .1636,1686. .1873,1970. .2090,
2137. .2316))
/ gene="SPAC1d4.02c"
complement(join(1088. .1636,1686. .1873,1970. .2090,
2137. .2316))
/ gene="SPAC1d4.02c"
/ note="SPAC1d4.02c, len:345, SIMILARITY:Rattus norvegicus,
035254, golgi peripheral membrane protein p65., (451 aa),
fasta scores: opt: 493, E():4.7e-20, (34.4% identity in
305 aa)"
/ codon_start=1
/ label="SPAC1d4.02c
/ product="putative golgi peripheral membrane protein"
/ protein_id="CAB56126.1"
/ db_xref="GI:5912525"
/ translation="MEGGLKNTFKKSEALAGIHRESDESGPRVLKVENDKAYNAR
IESYDFITRVNGLNLNGDPSFMALLRDSPEYTLLEVESLKQITRKVNIKINSDEK
IGVNLWASIAIPAVDAIWHILNVIDSPVARASLVPEYEDIVGTPEGMGKALSD
IESHLNRPRLRYLYNHVROSTROVTVVNRHMGNGAIGCGVGHVLRPLPAPLSGPP
PQSGDLYVSNMGLGSDPKVQSEPTENFLPTPEPKIASANAGSSMEISIPHYORK
KSHKGA10DSS10SYLDEEEKLSRELDHKTAKASSTWDQTPPLPPPPVAVNSTNE
SAPQNELVYKN"
misc_feature
complement(1637. .1651)
/ note="splice branch and acceptor sequence,
ctacaattacaag"
misc_feature
complement(1680. .1685)
/ note="splice donor sequence, gtagtgt"
misc_feature
complement(1874. .1886)
/ note="splice branch and acceptor sequence, ctacgactacaag"
misc_feature
complement(1964. .1969)
/ note="splice donor sequence, gtaagt"
misc_feature
complement(2091. .2108)
/ note="ctaaaggtttctacaag, splice branch and acceptor"
misc_feature
complement(2131. .2136)
/ note="gtataga, splice donor sequence"
gene
complement(join(2574. .2820,2907. .4201))
/ gene="SPAC1d4.03c"
complement(join(2574. .2820,2907. .4201))
/ gene="SPAC1d4.03c"
/ note="SPAC1d4.03c, len:513, SIMILARITY:Fugu rubripes,
O9YGN1, sand protein., (520 aa), fasta scores: opt: 609,
E():7.2e-32, (28.5% identity in 407 aa)"
/ codon_start=1
/ label="SPAC1d4.03c
/ product="hypothetical protein"
/ protein_id="CAA93212.1"
/ db_xref="GI:1177336"
/ db_xref="SWISS-PROT:Q10150"
/ translation="MEPTSEHSIKEEVNDNVHRSHESDEGSLNPGVNLAAAPSV
SEDDQEVSRSTPELRSHVENVEQLDILDNSSPLNVSTSVSSSNNTAVDEIKKL
SLSEFLAKOKRTYLRIPSSSGKPVSNIVDDISPEPTVGALOAIISSFEVSKELTSE
STESNVIVLSKNPILYLVGVSPTTSAVILISPLNLCOIILGTAKAMQOLINR
PNNDRLRLISNQFLKEICDOLNDYFLVPTLNKISPLPLRSSFRDLSOLLREPTK
SLAFTFAIRGLVCWAKKRLHLHANDLYLFLSTRTQSFNDSMEHWAPVCPETN
PDAYIYISYFLCKDYLVLINGSSESGVFEMQSVKCAOEI0DHGWLKLIYCEED
RTTPRNPGSPCISITLYFLYSKRSQFPPGYSFPNPNTRTLVLAIVSLDQAHKKN
SFSINMTVHSSLLFTWSTASDFHCIAANVTSSQLLIANVNLKLRWLRRENLFIO
TNLSF"
misc_feature
complement(2821. .2836)
/ note="splice branch and acceptor sequence,
ctaactatgttttta g"
misc_feature
complement(2901. .2906)
/ note="splice donor sequence, gtagtgt"
gene
join(5400. .5402,5439. .5570,5618. .7066)
/ gene="SPAC1d4.04"
/ note="cctc2"
/ note="SPAC1d4.04"
join(5400. .5402,5439. .5570,5618. .7066)
/ gene="SPAC1d4.04"
/ note="SPAC1d4.04"
/ note="SPAC1d4.04, len:527"
/ codon_start=1

```

```

misc_feature
5425. .5438
/ note="splice branch and acceptor sequence,
tactaacgataag"
misc_feature
join(5529. .5570,5618. .7045)
/ gene="SPAC1d4.04"
/ note="Match to PF00118 cpn60_TCP1, TCP-1/cpn60 chaperonin
family score 590.77"
5571. .5576
/ note="splice donor sequence, gtaagt"
5604. .5617
/ note="splice branch and acceptor sequence,
ttaacgataatag"
misc_feature
5741. .5767
/ gene="SPAC1d4.04"
/ note="P500995 Chaperonins TCP-1 signature 3"
complement(join(7282. .8366,8419. .8497))
/ gene="SPAC1d4.05c"
complement(join(7282. .8366,8419. .8497))
/ gene="SPAC1d4.05c"
/ note="SPAC1d4.05c"
/ note="SPAC1d4.05c, len:387, LOW SIMILARITY:Mus musculus,
Q920U0, polytropic murine leukemia virus receptor syg1.,
(695 aa), fasta scores: opt: 261, E():1.5e-10, (25.1%
identity in 398 aa) also shows low similarity to,
Saccharomyces cerevisiae, ERD1.YEAST, ERD1 protein., (362
aa), fasta scores: opt: 187, E(): 1e-05, (23.6% identity
in 382 aa)
SPAC1d4.05c, len:387, SIMILARITY:Mus musculus, Q920U0,
polytropic murine leukemia virus receptor syg1., (695 aa),
fasta scores: opt: 261, E():1.5e-10, (25.1% identity in
398 aa)"
/ codon_start=1
/ label="SPAC1d4.05c
/ product="hypothetical protein"
/ protein_id="CAA93214.1"
/ db_xref="GI:1177338"
/ db_xref="SWISS-PROT:Q10151"
/ translation="MDLEVEVEFPLHRLALPFRIGLVIVGTWLMSCYHILYILNRY
QPLSPNPGSLNSKMYHLQIPISNRTDLEENEFANILVSPDFRAGYCFALISL
SMATGFLFLFKTKTGIDGIGLYSHPIYPLVLTAVILIVPEFPRWYSSQGLRKSII
RVFLFPADEFSPGKDFVISEIFSYAKALQDFYIFCGVQGHISKFTLRDLCDGT
FVFLVLAAPVITVALIIOCINHGASRRHTEFIDMLSKHATAPVILASIIHAKQF
KFTTSGHGVLFWMITLSALSSAVTYLMOVEDITNLRPFRKSIINKRPFMFLYAG
CFITFLIRVITWSMKLHPRHLHOFHREMGHISFEMLLELRRLMLFFHLDAISS"
complement(8367. .8389)
/ note="splice branch and acceptor sequence,

```

Query Match 7.1%; Score 71.4; DB 15; Length 31201;
Best Local Similarity 54.9%; Pred. No. 7e-09;
Matches 141; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```

Oy 740 ttgcgtgttcgcaacatgttgcgaagagttcagaagctgcgtgcgcgcactgcga 799
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16446 TTCTTATTAAAGAAATTGATTAGTANCTTTTGACAGCGTTCCTGCACATGACTGCC 16505
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 800 gcatgcgtttgcgcgtgcgcgtgcacacctcgcgtgcgaagtgcgtctatccgcgga 859
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16506 GTATTCAATATTCAGATGAGATTACTTAATTCAGTGGTGGAGAGATCGCATCTTCGGAGA 16565
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 860 cacatgcacctgcgcgcgttcgcgcgtgcgaagagatgcacccttcgatacatga 919
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 1656 TACTGACCTCTTGAGAGATTGCTAACATTCGCAAGACTCACTTATTAATTCACGA 16625
Oy 920 ACCGACCCCGGAGAGATGCTTGGAGAGAGAACAGTGGGAGAGAGACAGCAGCAGC 979
Db 16626 ACCTACTTTGGATGATCTATCGAAGAAATTCGAATTAATAAAGACACAGACTTACTTC 16685
Oy 980 ccaagccatcagcgtg 996
Db 16686 TGAAGCTTTAGAAAGTGG 16702

RESULT 12
AC017383 115873 bp DNA HTG 09-DEC-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces
ACCESSION AC017383
VERSION AC017383.1 GI:6553603
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 115873)
AUTHORS Adams, M. and Venter, J.C.
TITLE Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
JOURNAL Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210513 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..115873
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 31582 a 25315 c 25855 g 33121 t
ORIGIN

Query Match 7.1%; Score 70.6; DB 64; Length 115873;
Best Local Similarity 49.2%; Pred No. 1 le-08;
Matches 346; Conservative 0; Mismatches 329; Indels 28; Gaps 5;

Oy 299 taccagaatacatcttccttggaacaggtctgcccacccgatgaagaatcgaatgctc 358
Db 63659 TACCGAAGATTAATTTCTTGGGACAGGCTCCTCGATTCCTCAATTAAGACGGCAACGTG 63718
Oy 359 agtgcacacatcgtcaacaagaagcccgacacgctctctgctactgactgactggtggagc 418
Db 63719 ACCTCCATTCATGATGAGACTGCATTAATGATGCTTGTGATGATGGAAGAAGA 63778
Oy 419 aatttggcagcgtgctgcgtcattacaggaacag---gtgacagaggtctcggcacc 475
Db 63779 ACTTTATGGCAATTTGATGAGATATATGACACGAGAAAGGCAAGTTGATTCCTGCCAG 63838
Oy 476 ctgagctgctgttctgtgtccacactggaacgacatcaccaacagggcttgcacagat 535
Db 63839 CTACAGGGCATTTATGTATGCCA-TTTGCATGCCGACACACATTTGATTTGCTGT 63897
Oy 536 ctgtctgcagagagaagcgctgtgcatcttgggaaagccgctcacccttctgtgt 595
Db 63898 GCTCCGGGAAAG-----GAGGCAATTAACCTTAGGGGAGACCCACTTATTC 63945
Oy 596 ggttcccccaacacagctcaagcctgtgtccagcagatcaccaacacagtgccagaggt 655
Db 63946 GCTGCTCTCTCCCAAAATTAAGACCTTGTT---GGAGTTCTACAAATCGACAAATTAAGAAC 64002
Oy 656 cctgacacacatcagatgattcctgcgcaaatgcttcaggaagggcgtgagatctccag 715

```

```

Db 64003 CGTAGAAGATGCGTATCTAGTGGGCAACGCGTAACCTCTAGTAGCCCTAGCGCG 64062
Oy 716 tccctcagcaggaagaatgatcatcagttcgtgtgtggaacatgtattggaagattca 775
Db 64063 TGAACAGTTGAACGCTTAGAATTAACGTCATATCCACTCTCCTAGTTAGCCACTGCC 64122
Oy 776 gacctgtcgtgctgagcgaactgcaagcatgcttgcgtgtgcgtgtgcacactctg 835
Db 64123 CAACTCTTTGGGAATTAAGCGCTTAACCTGTGGC-----GCAAGACAAATAGGAACC 64173
Oy 836 ctggaagatggtctatctccgaggaacacatccctcgtcagcagctgtgtccgattggaa 895
Db 64174 CGTCAAGATCACTTACAGCGGCGACACTATGCCATGTCAAGATCTGATCTGGCGCG 64233
Oy 896 agatgcaccctccctgatcatgaaagccaccttgaagaatggttggaaaggaagcagt 955
Db 64234 CGACTCCACCGTTCTTATTCAGAGAGCGCAATGGAGGATGACTTGGAGAGAGAGCGCG 64293
Oy 956 ggaagaacacacagacagcaacgctcccaagcactcagcgtgag 998
Db 64294 GTTAAAGACGACAGACCGCTTCGAGCGCTATTCAAGCAGGGG 64336

RESULT 13
AC007417 175682 bp DNA INV 02-MAR-2001
LOCUS Drosophila melanogaster, chromosome 2R, region 47A-47B, BAC clone
DEFINITION BACR48F07, complete sequence.
ACCESSION AC007417
VERSION AC007417.4 GI:13184056
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 175682)
AUTHORS Celinker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Fartan, D.,
Ferreira, S., Frishe, E., Galle, R.F., Gary, N.S., George, R.A.,
Gonzalez, M., Houck, R.A., Hoskins, R.A., Hostin, D., Howland, T.J.,
Izbagyan, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunco, J.,
Pacheb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phoumenavong, S., Piltman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
TITLE Sequencing of Drosophila chromosome 2R, region 47A-47B
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175682)
AUTHORS Celinker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazer, R.G.,
Buenhofer, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Fartan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummst, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomutan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacheb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Shlir, E.,
Svirska, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Mar 2, 2001 this sequence version replaced gl:5670592.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.

```


*	14599	18837:	contig of 4239 bp in length
*	16838	18917:	gap of unknown length
*	18918	24324:	contig of 5407 bp in length
*	24325	24404:	gap of unknown length
*	24405	29984:	contig of 5580 bp in length
*	29985	30064:	gap of unknown length
*	30065	35794:	contig of 5730 bp in length
*	35795	35874:	gap of unknown length
*	35875	44808:	contig of 8934 bp in length
*	44809	58222:	contig of 13334 bp in length
*	58223	58302:	gap of unknown length
*	58303	78322:	contig of 20020 bp in length
*	78323	78402:	gap of unknown length
*	78403	122249:	contig of 43847 bp in length
*	122250	122339:	gap of unknown length
*	122339	179102:	contig of 56773 bp in length
*	179103	179182:	gap of unknown length
*	179183	179407:	contig of 225 bp in length
*	179408	179487:	gap of unknown length
*	179488	180592:	contig of 1105 bp in length
*	180593	180672:	gap of unknown length
*	180673	181383:	contig of 711 bp in length
*	181384	181463:	gap of unknown length
*	181464	182092:	contig of 629 bp in length
*	182093	182172:	gap of unknown length
*	182173	183480:	contig of 1308 bp in length
*	183481	183560:	gap of unknown length
*	183561	183771:	contig of 211 bp in length
*	183772	183851:	gap of unknown length
*	183852	184058:	contig of 207 bp in length
*	184059	184138:	gap of unknown length
*	184139	185035:	contig of 897 bp in length
*	185036	185115:	gap of unknown length
*	185116	185846:	contig of 731 bp in length
*	185847	185926:	gap of unknown length
*	185927	186560:	contig of 634 bp in length
*	186561	186640:	gap of unknown length
*	186641	187253:	contig of 613 bp in length
*	187254	187333:	gap of unknown length
*	187334	187611:	contig of 278 bp in length
*	187612	187691:	gap of unknown length
*	187692	188409:	contig of 718 bp in length
*	188410	188489:	gap of unknown length
*	188490	189196:	contig of 707 bp in length
*	189197	189276:	gap of unknown length
*	189277	189969:	contig of 693 bp in length
*	189970	190049:	gap of unknown length
*	190050	190574:	contig of 525 bp in length.

FEATURES	SOURCE
Location/Qualifiers	1. 190574
/organism="Drosophila melanogaster"	
/strain="y: cn bw sp"	
/db_xref="taxon:7227"	
/chromosome="2"	
/map="47A-47B"	
/clone="BACR19J16 (D626) RPCI-98 19.J.16"	
/clone_1lb="RPCI-98 (Roswell Park Cancer Institute	
Drosophila melanogaster BAC library, partial EcoRI in	
pBacc3.6"	
BASE COUNT	53370 a 40358 g 54299 t 2821 others
ORIGIN	39726 c

Query Match	7.1%;	Score 70.6;	DB 60;	Length 190574;
Best Local Similarity	49.2%;	Pred. No. 1e-08;		
Matches 346;	Conservative 0;	Mismatches 329;	Indels 28;	Gaps 5;

Qy 299 taccagaatacatcttccttggaacaggtcgcacatccgagtgaagattcgaaatgtc 358
||||| | |||| | ||||| || |||| | ||||| ||||| ||
Db 132812 TACCGAAGTAAATTTCTTGGGCACAGGCTCTGCATTCCCAATAAGACGGGAACGTG 132871

QY 359 agtgcacactgtcaacataagcccgacacgctctctgctaactgactgtgtgaggc 418

D	132872	AGCTCATTTCTGATCAGACGACGACATTCGCTACGTTTGTGGATTGTGGACAGGA	132931
Q	419	acaattggcagctgctgcccgaatcaggaacag--gtagacaaggctcctggcacc	475
D	132932	ACTTATGTGCAAATTGTACGACTATATGTGACACGAGAAAGGCGACAGTGTCTTGCCGAG	132991
Q	476	ctggcctgcctgcttcttggctccaccctggacagacatcaccaacagggcttccaagat	535
D	132992	CTACAGCGCAATTTATGTATCCCA-TTTGGATGCGCACACCACTTGGATTATTTGGTCT	133050
Q	536	ctgtctgcagagagaacgcgcttggcatctttggaaagacgcgcttcaaccttgcgtgt	595
D	133051	GCTCCGGGAAAG-----GAGGCAATTAAACCTAGGCGACCCACTTATCTCT	133098
Q	596	ggttgcgcccaaccagctaaagccttgcgtccagcagtcaccacaaccgttgcaggaggt	655
D	133099	GCTGTGCTCTCGCCCAATTGAAACTTTGGTT--GGAGTTCACAAATTCAGCAAAATGAAAC	133155
Q	656	cctgcacccacatcagatgatcttcctgcgcaaatgccttcgaaagggcctgagatccag	715
D	133156	CGTAGAAGATGGGTATACCTAGTGGGCAACGCTGAACCTAGTAGTACCCCTGAGCGG	133215
Q	716	tccctgcagctggaaagatttgatcagtttcgctctgtcttgcgaacatgtgatttggaaagattca	775
D	133216	TGAACCAAGTTTGAAACGCTTAGGAATTACGTCACATTCACACTCTCTTAGTAGTAGGCACTCGCC	133275
Q	776	gaacctgtctgtgtgacgcgacatgcgaacatbctgttctgcttgcctcgtgtgcacaccttbg	835
D	133276	CAACTCTTTGGGAATTAAGCCTAACTCTGCG-----GCMAAGACAAATTAAGGAAC	133326
Q	836	ctggaaagtgtctctattccgggggaaacacatgcctctgcgaggtcctgtgtccgaatgggaa	895
D	133327	CGTCAAGATCACTTACAGCGGGGACACTATWGCAATGTCAGATCTGGCCG	133386
Q	896	agatgcacacctctctgatatacgaagccaccccttggaaagatgttctggaaagagaagcagt	955
D	133387	CGACTCCACGTTCTTATTTCACGAGGCGCAATWGGAGATGACTTGGAGAGAGAGCGCG	133446
Q	956	ggaaaaagacacagcagcagcagctcccaagccatcagcgttggg	998
D	133447	GTTTAAAGACGACACACCGTGTGCGAGGCTATTTCAGCAGGGG	133489

RESULT	15
AEO003830/c	
LOCUS	
DEFINITION	AEO003830 261846 bp DNA INV 04-OCT-2000
ACCESSION	Drosophila melanogaster genomic scaffold 142000013386047 section 13 of 52, complete sequence.
VERSION	AEO003830 AE002787
KEYWORDS	AEO03830.2 GI:10727666
SOURCE	HTG.
ORGANISM	fruit fly. <i>Drosophila melanogaster</i> Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; <i>Drosophila</i> . 1 (bases 1 to 261846)
REFERENCE	Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
AUTHORS	

Aadams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gockayne, J., D'Amico, D., Akiyoshi, K., Lachy, J., Rosenblatt, M., Wang, S., Gish, W.B., Ziegler, W., Adams, J.R., Maniatis, P.G., Scherer, S.E., Il, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Vandeil, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazek, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, M.Klos, G.L., Abril, J.F., Abmayyan, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Bernier, B.P., Bhandari, D., Bolintakov, S., Borrova, D., Botchan, M.R., Bouck, J., Brockschtein, P., Brotlier, P., Burris, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, R., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,

Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Fertiera, S., Fleischmann, M., Gleser, K., Glodok, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegam, C., Jallali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Laekko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Mikhina, N.V., Mobarry, C., Morris, J., Mosher, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Paclob, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 261846)
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7303755.
location/Qualifiers
I. 261846
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
join(960..1069,1135..1237,1290..1357,1416..1731,2658..2759,3801..4142,4200..4354,4727..4835,6101..6289,6499..6941,6999..7841,7921..8315)
/gene="PIK"
/product="CT13302"
/db_xref="FLYBASE:Fban0004001"
/db_xref="FLYBASE:Fbgn0003071"
<960..>8315
/gene="PIK"
/note="CG4001"
/map="46E3-46E6"
/db_xref="FLYBASE:Fban0004001"
/db_xref="FLYBASE:Fbgn0003071"
join(1026..1069,1135..1237,1290..1357,1416..1731,2658..2759,3801..4142,4200..4354,4727..4835,6101..6289,6499..6941,6999..7841,7921..8059)
/gene="PIK"
/note="PIK gene product [alt 1]"
/codon_start=1
/db_xref="FLYBASE:Fban0004001"
/db_xref="FLYBASE:Fbgn0003071"
/protein_id="AAFS8840.1"
/db_xref="GI:7303792"
/translation="MHSIKRPFVTKLPIFLFINGRIPICRHFGPTTFLTSNKP
PIROKLTFFNIGICTRSHHLCPCPDISGNTLLSVKRCRHCILKSDSGDKNDSP
GEKNIDKSAQRCQKPIINLHGFNLAVSENAVKRKSAPKRCGSDVLRK
LRTMODVIDFVHGVKPKDKGLAVFTSGDSQANNAVRACVRNALYLGKVFIRG
YQGVGSGDCTIQANNAVSIIHIGRTIIGSRCOFPEROGRLKANNLLDGIITN
LVYIGGSGSLQANNAVSIIHIGRTIIGSRCOFPEROGRLKANNLLDGIITN
GTMITGDTALRIIEAIDAISSASTASHORTFTMEYMGHGLIAGVLACADEFI
FTEMPKVDMPDRILCSQLQOERSAGRLNIVYABGAMRBEHPITAEVKKYIDR
LKHDAITVLGHVQRGNSPASFDTILLCRMGAETTLALMEATKDSVPVVISLGNQAV
RVLPLMCEVERTQAVAKAMAEKRNADAVKLGSRSEFTALMKLTRLKPKENEDAG
KGIEGYRLAVMHIGAPACGMNAVRSVRNAYIRGVGVYIGINDVEGLIAGNVELM
SDVSGWVGGAIVLGTIRLTPEKFKETIARLKEPKIOGLLIGGFESYHAAQIDAO
RDWYPOCFIIVIPITISNNVGTGERSLGDGILNITEITCIRIRISAGCTKRIVY

ITMTGYCGYLATIAGAGADAAYIEEFKSIKDLQODVYHNAKMAEVSRLILR
NEASENYSSTDPIYLYSEEGKGLFTCRNMILHMOGGSPTEPDRMGKMAKCD
WLAQIKANIDANGVYCKSPDRTLLGILYSROYRSPVLVDIAETNFDRIKQKQW
LRPLRLITIAKDSAEERGGITTYEEDDTAVA"
join(2152..2759,3801..4142,4200..4354,4727..4835,6101..6289,6499..6941,6999..7841,7921..8315)
/gene="PIK"
/product="CT13275"
/db_xref="FLYBASE:Fban0004001"
/db_xref="FLYBASE:Fbgn0003071"
join(2613..2759,3801..4142,4200..4354,4727..4835,6101..6289,6499..6941,6999..7841,7921..8059)
/gene="PIK"
/note="PIK gene product [alt 2]"
/codon_start=1
/db_xref="FLYBASE:Fban0004001"
/db_xref="FLYBASE:Fbgn0003071"
/protein_id="AAFS8841.1"
/db_xref="GI:7303793"
/translation="MNSINORFLARSGOKDGLAVFTSGDSQANNAVRACVRNAI
YICGYFIREGVOGWDCDCTQANMAVSSTIHRGTTISARCODRERQRLK
AANNLIDRGITNLVIGDSLTGANLFEWMSLDELKNTTTEQOKRFLVLT
VGLVSTIDNFCGTDMTIGDTALRIIEAIDAISSASTASHORTFTMEYMGHGLI
LVGLACADPFIPEMPKVDMPDRILCSQLQOERSAGRLNIVYABGAMRBEHPIT
TAEDVKKVIDERLKHDAITVLGHVQRGNSPASFDTILLCRMGAETTLALMEATKDS
VPVVISLGNQAVRPLMCEVERTQAVAKAMAEKRNADAVKLGSRSEFTALMKLT
LRPKENEDAGKIEGYRLAVMHIGAPACGMNAVRSVRNAYIRBDVYIGINDVE
GLIAGNRELWMSDVSQVQGAIVLTCTKTLPEKFKETIARLKEPKIOGLLIGSF
ESYHAAQIDAOVDWYPOCFIIVIPITISNNVGTGERSLGDGILNITEITCIRIR
ISAGQTKRVRVLETMTGCGYLATIAGAGADAAYIEEFKSIKDLQODVYHNAK
MAEVSRLILRNEKASENYSSTDPIYLYSEEGKGLFTCRNMILHMOGGSPTEPDR
MGTKMAKCVDMIAQIKANIDANGVYCKSPDRTLLGILYSROYRSPVLVDIAET
NFDRIKQKQWLRPLRLITIAKDSAEERGGITTYEEDDTAVA"
join(8915..9052,9216..9726,10663..10911,10976..11174,11235..11542,11596..12210)
/gene="CG11866"
/product="CT37014"
/db_xref="FLYBASE:Fban0011866"
/db_xref="FLYBASE:Fbgn0033486"
<8915..>12210
/gene="CG11866"
/map="46E4-46E6"
/db_xref="FLYBASE:Fban0011866"
/db_xref="FLYBASE:Fbgn0033486"
join(9286..9726,10663..10911,10976..11174,11235..11542,11596..12210)
/gene="CG11866"
/note="CG11866 gene product"
/codon_start=1
/db_xref="FLYBASE:Fban0011866"
/db_xref="FLYBASE:Fbgn0033486"
/protein_id="AAFS8839.1"
/db_xref="GI:7303791"
/translation="MVSTROMSGSGTSSAGEVSPFGAGSGASGTGMASGAVSSG
SGTGVQEPPESTANRSVLRRITSYGNOHSSYRVVSAAGSLAESAVRILHSYTMO
MQORPPTATTTTTSNPSIYDPEINLEIKTLISVDYKRVSNLRVLSHRANDICMAH
NTAATKQIKTTLTSRPAIKASMPRESHRNRNPLACDILITETYMNLVDLWMSGH
IERHCCFEPGAILDEVQALINYSITPRQRPYRVNDELFDLSTMAEYFKQIEAT
LPGIAFNKDFYLTPTTKRPTLAISDLSDSASNSPPONHMYLRKQIRIKQGMKY
NNOLSVLRETITRCRKAAOSKOLASOINISQOCTLEYARLEALNKENKMSER
KFTSLLOELKCKTELOFWRSKSPAIYAVCSOQKAPVVPDPDPAALNOGVDPH
FIIINEVDASDVSVGELEKSPASDSTTAKLALAYATAKNLKRKPSSESQALATS
TSRAEVAQSQSPMAKNGAAGAGYSFKLFGHHSQVIVSPSONLGAATT
DNAAGSEALAEPEAKRKYVAAKGYVTPGKRS"
complement(join(<12115..12937,12999..13642,13694..13708))
/gene="CG17753"
/product="CT11457"
/db_xref="FLYBASE:Fban0017753"
/db_xref="FLYBASE:Fbgn0033487"
complement(<12115..>13708)
/gene="CG17753"
/map="46E4-46E6"
/db_xref="FLYBASE:Fban0017753"

CDS

complement[[join](#)(12802, .12937,12999, .13642,13694, .13708))

/db_xref="FLYBASE:FBgn0033487"

/gene="CG17753"

/note="CG17753 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan00118753"

/db_xref="FLYBASE:FBgn0033487"

/protein_id="AA58838.1"

/db_xref="GI:7303790"

/translation="MSKIEFAVMRRGDESYAGALRSALDVGVEIDTOEGRVIT
OTORWMSVIOKIEATGAPRAVLSEGSQASALINTGTSVDPTPIOGVRRFTITAD
KKPGVWVGIOGLSPGLHGHIHESDTSAGCSVGEHYNPPOSPHSGPAAETTRD
AGDGNITADNGRATRFREVDPLVEWDTIGRAVVL7ANADDLGRGNGQSLIDGNSG
EIRAGIIARSAGLLENFKRICACDGYTLWDERKKPLACKERSQKI."

join(<18636, .18761,19765, .19876,19968, .20213,22806, .22878,
22956, .23344,23489, .23633,23934, .24407,25026, .>25202)

/gene="CG11867"

/product="CT37016"

/db_xref="FLYBASE:FBan0011867"

/db_xref="FLYBASE:FBgn0033488"

<18636, .>25202

/gene="CG11867"

/map="4GE4-46E8"

/db_xref="FLYBASE:FBan0011867"

/db_xref="FLYBASE:FBgn0033488"

join(18636, .18761,19765, .19876,19968, .20213,22806, .22878,
22956, .23344,23489, .23633,23934, .24407,25026, .25202)

CDS

gene

Query Match	7.1%;	Score 70.6;	DB 5;	Length 261846;
Best Local Similarity	49.2%;	Pred. No. 1e-08;		
Matches 346;	Conservative 0;	Mismatches 329;	Indels 28;	Gaps 5;

Oy 239 taaccagaatcaatcttctcttggaacagggtctgcaatcccgatgaagatctgaatgtc 358
 Db 199291 TACCAGGAATTAATTTCTTGGGCAAGGCGCTCGAATCCCAATAAGACGCAAAAGTG 1992322
 Oy 359 agtgcacaactgtcaacataagcccgacaacgctctgtctactgagactgtgtgaagc 418
 Db 199231 AGCTCCATCTTGATCAGGACACGCAATTGATGCCCTACGTGTTGTTGGATTGTGGAGAAGCA 1991722
 Oy 419 acatttggcagcgtgtgcgcgtcattacggaagcag--gtgacaggtctcttggacc 475
 Db 199171 ACTTATGGTCAAAATGTATACGACTATATGGACACGAGAAAGGGCAGTTGATTCTTCGGCAG 1991122
 Oy 476 ctgctgctgtgtgtgtgtgtgcctaccctgcgcagcagatcaccaacagcgcttgcgaagtat 535
 Db 199111 CTACAGGGGATTTATGTAATCCCA--TTTGCAATGCGGACCCACACTTGGAATGATGGTCT 1990533
 Oy 536 ctgtctgcagagaagaacgcgcctctgtgacctcttggagaagcgcctcacctcttgcgt 595
 Db 199052 GCTCCGGGAAG-----GAGGCAATTMAAACATGAGGACAGACCCTATTCT 199005
 Oy 596 ggttgcccccaaccagctcaaaagcctgtgtctcagcagtatccacaaccagtgcagaaggt 655
 Db 199004 GCTGGCTCTCCGCCCAATTGACCTTGTT---GGACTCTTCAAAATGCAAAATAGAAAC 198948
 Oy 656 cctgcacacatcagtatgtatctctgcgaatgcttcacagaaaggcgctgagctctccag 715
 Db 198947 CGTAGAAGATCCGTAATCTCTAGTGGGCAACGGTGAACCTCTTAGCTAATGCTCCCTGAGCG 198888
 Oy 716 tcttcagagtggaaagatltgacagtlcgcgtgttgcgacatgtgatttggaaagattca 775
 Db 198887 TGACAAGTTGAACGTGTAGGAATTAAGTCCATATTCACACTGCGCTAATTAGGCACTGCC 198828
 Oy 776 gactgtctgtgtgcgcacgtgcgaagcatgcttgtgtctgtgcgtctgtgcacacctg 835
 Db 198827 CAACCTTTTCGGAATTAAGCCCTAACCTTCGGC-----GCAAAAGCAAAATAGCGAACC 198777
 Oy 836 ctggaagtgctatctccgcgagacacatgtccctcgagagctgtgtcgcgattggaa 895
 Db 198776 CGTCAGATCTTACACAGCGCGCACATATGCCATGTGCAGAGATCTGATCATCTGGGCG 198717
 Oy 896 agatgcaaccctctctgtatcagtcgaacaccttgaagatgtgttggaaagagaaagcagt 955

Db 198716 CGACTCACCCGCTTCTATTATTCAGAGGCGACAAATGAGAGATGACTTGGAGGAGGAGCGC 198657

QY 956 ggaagaagacacacagcacacagctcccaagcatcaagctggyg 998

Db 198656 GTTAAAGACGCACAGCACCGCTGTGTGCAGGCGATTATTCAGACAGGG 198614

Search completed: October 28, 2001, 21:51:39
Job time: 6372 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 21:04:27 ; Search time 100.36 Seconds
(without alignments)
6262.749 Million cell updates/sec

Title: SEGLINS_COPY_1140_2140
Perfect score: 1001
Sequence: 1 tcgcagcccaacgaattcaaa.....aagccatcagcgttgaggatc 1001

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
N_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	989	98.8	2546	21	AACT6445
2	989	98.8	2958	21	AAA58453
3	984.2	98.3	2478	21	AAA52810
4	136.6	13.6	2664	21	AAA60207
5	70	7.0	1617	21	AAAC34173
6	70	7.0	2853	21	AAAC49965
7	41.6	4.2	464	21	AAAC38171
8	37.6	3.8	521	21	AAAC04378
9	37	3.7	2517	22	AAAC89621
10	36.4	3.6	10732	21	AAA10594
11	36	3.6	1409	21	AAAC2402

C	12	35.8	3.6	19053	21	AAAS9146
C	13	35.4	3.5	366	20	AAAB9531
C	14	35.2	3.5	1716	22	AAAF26315
C	15	35	3.5	3474	14	AAQ39018
C	16	35	3.5	6901	18	AAAT78853
C	17	35	3.5	8854	20	AAAZ20086
C	18	34.8	3.5	4173	21	AAAS2899
C	19	34.8	3.5	4200	21	AAAS2900
C	20	34.8	3.5	4266	21	AAAS2902
C	21	34.8	3.5	4269	21	AAAS2901
C	22	34.2	3.4	2024	11	AAAO3143
C	23	34	3.4	300	20	AAZ4375
C	24	34	3.4	2022	21	AAZ49455
C	25	33.8	3.4	928	21	AACT7016
C	26	33.4	3.3	1032	21	AAAF07546
C	27	33.4	3.3	1128	20	AAZ06453
C	28	33.4	3.3	3474	18	AAAT74012
C	29	33.4	3.3	3474	19	AAAV16191
C	30	33.4	3.3	3508	14	AAO39020
C	31	33.4	3.3	3508	18	AAAT74013
C	32	33.4	3.3	3508	19	AAV16192
C	33	33.4	3.3	3840	10	AAAN2365
C	34	33.2	3.3	1791	19	AAV59723
C	35	33.2	3.3	1800	22	AAAF61102
C	36	33	3.3	569	21	AAAC56065
C	37	33	3.3	3483	21	AAAG19765
C	38	33	3.3	43804	18	AAAT66375
C	39	33	3.3	43804	20	AAAT2690
C	40	32.8	3.3	696	21	AAAS50885
C	41	32.8	3.3	1401	21	AAAZ23803
C	42	32.8	3.3	1872	11	AAAO6828
C	43	32.8	3.3	2385	18	AAAT72725
C	44	32.8	3.3	2871	21	AAZ50586
C	45	32.8	3.3	3600	21	AAAB9736

ALIGNMENTS

RESULT	1
ID	AACT6445
XX	AACT6445 standard; cDNA: 2546 BP.
XX	
AC	AACT6445:
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human ORFX ORF2000 polynucleotide sequence SEQ ID NO:3999.
XX	
XX	Human: open reading frame; ORFX: detection; cytosstatic; hepatotropic; vulnerrary; antiporiatic; antipartinsonian; noctropic; neuroprotective; anticonvulsant; osteopathic; antilarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antinflammatory; antitviral; antibacterial; antifungal; antirheumatic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antinflammatory disease; coagulation; thrombosis; contraceptive; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.

Nucleotide sequenc
EST clone CP314.
Pseudomonas sp lip
Sequence of pure m
Human lecithin-cho
DNA encoding chime
Murine JNK3 bindin
Murine JNK3 bindin
Murine JNK3 bindin
Murine JNK3 bindin
Murine JNK3 bindin
Nucleotide sequenc
Human gene express
Human Voltage-gate
Human ORFX ORF2571
Fusarium venenatum
cDNA of Baboon Gro
Maize optimised-B.
Maize optimised-DN
Sequence of full l
Full length maize
Maize optimised DN
Sequence of maize
Human secreted pro
P. putida KT2440-a
Eucalyptus grandis
Mouse trophinin DN
Chicken embryo let
Complete genome se
A. sulcata humanis
Human Kv6.2 cDNA
Extracellular port
Her2-GM-CSF immuno
DC8scFv-erbB2EC fu
Human HER-2/neu co

PR 02-APR-1999: 99US-0127636.
PR 05-APR-1999: 99US-0127728.
PR 30-MAR-2000: 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI: 2000-602362/57.
DR P-PSDB: AAB42236.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX
PS Claim 5; Page 3179-3180; 5507pp; English.
XX
XX AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
CC osteoparitic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antidiarrhetic; antibacterial; antiviral; antifungal; antineumatic;
CC antihypoid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX
SQ Sequence 2546 BP: 652 A; 643 C; 686 G; 564 T; 1 other;

Query Match 98.8%; Score 989; DB 21; Length 2546;
Best Local Similarity 99.9%; Pred. No. 1.1e-266;
Matches 1000; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 tcgcagccacaagaattcaaacccagctcaacctcattccacccgggaattctccctgct 60
DB 738 tcgcagccacaagaattcaaacccagctcaacctcattccacccgggaattctccctgct 797
QY 61 caccagcttcgcgtgtaagaaggagcccaacctcagtggtgccatgtttcaagggtga 120
DB 798 caccagcttcgcgtgtaagaaggagcccaacctcagtggtgccatgtttcaagggtga 857
QY 121 atgcctctctaaatlacagcttcggtcccaaggaggagtgagcagagatccattatcac 180
DB 858 atgcctctctaaatlacagcttcggtcccaaggaggagtgagcagagatccattatcac 917
QY 181 ttgcaatcctgaggaattcattgttaggcgctgcagcttcccaacttcagcagagcgt 240
DB 918 ttgcaatcctgaggaattcattgttaggcgctgcagcttcccaacttcagcagagcgt 977
QY 241 gcagaggttaagagaggtgcgcagagagcccgcccgagcagaagaagaattcagta 300
DB 978 gcagaggttaagagaggtgcgcagagagcccgcccgagcagaagaagaattcagta 1037
QY 301 cccagaatcatcttcccttgaaacagagtgccatcccgatgaagaattcgaattgcag 360
DB 1038 cccagaatcatcttcccttgaaacagagtgccatcccgatgaagaattcgaattgcag 1097
QY 361 tgcacaactgttaacaataagcccgacaagctctgtctactgtgactgtgttgaggcac 420
DB 1098 tgcacaactgttaacaataagcccgacaagctctgtctactgtgactgtgttgaggcac 1157

QY 421 atttgggcaagctgtgcccgtcattacggagagccaggtgagcagggtctctgtgcaacctgac 480
DB 1158 atttgggcaagctgtgcccgtcattacggagagccaggtgagcagggtctctgtgcaacctgac 1217
QY 481 tgcgtgtttgtgttcccaacttgcacgcagatcacccacacggcgttgcgaagtatcttc 540
DB 1218 tgcgtgtttgtgttcccaacttgcacgcagatcacccacacggcgttgcgaagtatcttc 1276
QY 541 tgcagagagaagcgcgcttggcatctttgggaaagccgcttcaccccttctgtgtgttg 600
DB 1277 tgcagagagaagcgcgcttggcatctttgggaaagccgcttcaccccttctgtgtgttg 1336
QY 601 cccccaacacgctcaaaagccttgcctcagcagtaaccacaacagttgcagaggtctctgc 660
DB 1337 cccccaacacgctcaaaagccttgcctcagcagtaaccacaacagttgcagaggtctctgc 1396
QY 661 accacatcatatgatattccttgcgaatgcttcaggaagggtcgtgatctccagctcgt 720
DB 1397 accacatcatatgatattccttgcgaatgcttcaggaagggtcgtgatctccagctcgt 1456
QY 721 cagtggaagaattgatcagttcgctgttgcgaatattgatattggaagatttcaagcct 780
DB 1457 cagtggaagaattgatcagttcgctgttgcgaatattgatattggaagatttcaagcct 1516
QY 781 gtcgtgtgcgcaacttgcgaacatgcttgcgttgcgtgcgtgcgcacaccttgcctgga 840
DB 1517 gtcgtgtgcgcaacttgcgaacatgcttgcgttgcgtgcgtgcgcacaccttgcctgga 1576
QY 841 aagtgtctatctccggggaacacatgcccctgcagagctctgttcggatgtgggaaagatg 900
DB 1577 aagtgtctatctccggggaacacatgcccctgcagagctctgttcggatgtgggaaagatg 1636
QY 901 ccacccctcgtatataatgaagccaccttggaaagattgttggaaagaaagcagtgtaa 960
DB 1637 ccacccctcgtatataatgaagccaccttggaaagattgttggaaagaaagcagtgtaa 1696
QY 961 agacacacagcacaaagctcccaagccatcagcgtggggatg 1001
DB 1697 agacacacagcacaaagctcccaagccatcagcgtggggatg 1737

RESULT 2
AAA58453
ID AAA58453 standard; cDNA; 2958 BP.
XX
XX AAA58453:
XX
XX 07-DEC-2000 (first entry)
XX
DE Human prostate cancer predisposing gene HPC2 coding sequence.
DE
XX
XX Human: prostate cancer predisposing gene; HPC2; chromosome 17p.
KW gene therapy; peptide therapy; drug design; ss.
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT CDS 51..2531
FT /tag= a
FT /product= "HPC2"
XX
PN WO200027864-A1.
PD 18-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26055.
PF
XX
XX 06-NOV-1998; 98US-0107468.
PR
XX (MYRI-) MYRIAD GENETICS INC.
PA
XX TAVTigian SV, Teng DHF, Simard J, Rommens JM;

XX MP1: 2000-376481/32.
DR P-PSDB: AAB07228.
XX
XX Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
XX
PS Claim 3; Page 98-100; 157pp; English.
XX
XX The present sequence is the coding sequence of the human prostate
CC cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC alleles of this gene cause a predisposition to cancer, particularly
CC prostate cancer. This gene and its protein can be used in peptide and
CC gene therapy for cancer sufferers, as well as being useful as diagnostic
CC tools (both for cancer sufferers and those with a predisposition to the
CC disease) and in the production of cancer drugs. This sequence was
CC isolated by cloning and sequencing the region of the genome which
CC appeared to cause a predisposition to prostate cancer.
XX
XX Sequence 2958 BP; 707 A; 805 C; 848 G; 598 T; 0 other;

Query Match	98.8%	Score 989	DB 21	Length 2958
Best Local Similarity	99.9%	Pred. No. 1.2e-265		
Matches 1000	Conservative 0	Mismatches 0	Indels 1	Gaps 1
OY 1	tcgcagccacaagatctcaaacccgagctcaactctcaactctcaaccccgagacatctccctgct	60		
Db 1190	tcgcagccacaagatctcaaaccccgagctcaactctcaaccccgagacatctccctgct	1249		
OY 61	caccgagtttcgcgtgtaagaaaggagagcccccctcagtggtccatggtctcaggtgta	120		
Db 1250	caccgagtttcgcgtgtaagaaaggagagcccccctcagtggtccatggtctcaggtgta	1309		
OY 121	atgcctctctcaagatcacccgctccgtcccaagagagatgtagcagaggtatgcatatcac	180		
Db 1310	atgcctctctcaagatcacccgctccgtcccaagagagatgtagcagaggtatgcatatcac	1369		
OY 181	ttgcgacatccctgagagatctcatatgattgtagagcggtgtagcgtctcccaacttccagcagagcgt	240		
Db 1370	ttgcgacatccctgagagatctcatatgattgtagagcggtgtagcgtctcccaacttccagcagagcgt	1429		
OY 241	gcagagagatcaagaaaggagatgtagcagagacgagcccaagccccaagagagaaagaagttcagta	300		
Db 1430	gcagagagatcaagaaaggagatgtagcagagacgagcccaagccccaagagagaaagaagttcagta	1489		
OY 301	cccagaatactatctctctcttggaacagaggtctgcatctccgattgtaagatctcgaaatgtagcag	360		
Db 1490	cccagaatactatctctctcttggaacagaggtctgcatctccgattgtaagatctcgaaatgtagcag	1549		
OY 361	tgccgacactgtgcaacaataaagcccggaacaggtctctgcatctgagacgtgtagatgtagtagggcagc	420		
Db 1550	tgccgacactgtgcaacaataaagcccggaacaggtctctgcatctgagacgtgtagatgtagtagggcagc	1609		
OY 421	atttgagcagctgtagtcggtcatctacgagaaaccaaggtgtagacaggtctctgtagcaccctgtagc	480		
Db 1610	atttgagcagctgtagtcggtcatctacgagaaaccaaggtgtagacaggtctctgtagcaccctgtagc	1669		
OY 481	tgctgtgttctgtgtcccaacctgtagcagcagagatcaaccaacagaggtctgcaagatctctgtagc	540		
Db 1670	tgctgtgttctgtgtcccaacct-gtagcagcagatcaaccaacagaggtctgcaagatctctgtagc	1728		
OY 541	tgcaagagaaagcgcgtctgtagcatctcttgtagaaagcgcgtctcaaccccttgtaggtggtg	600		
Db 1729	tgcaagagaaagcgcgtctgtagcatctcttgtagaaagcgcgtctcaaccccttgtaggtggtg	1788		
OY 601	cccccaacacagatctcaaaagctggtcttcacagatcaccaacaacagatgtagcaggaagtctctgtagc	660		
Db 1789	cccccaacacagatctcaaaagctggtcttcacagatcaccaacaacagatgtagcaggaagtctctgtagc	1848		
OY 661	accacatcagatgattctctgccaatagctcttcagagaagaggtgtagatctccagatctctgtagc	720		
Db 1849	accacatcagatgattctctgccaatagctcttcagagaagaggtgtagatctccagatctctgtagc	1908		

Qy	721	cafrvgsgaaagatlgatcatagctgcgtctgttcggaacatgabatcttggaagagtttcagact	780
Db	1909	cafrvgsgaaagatlgatcatagctgcgtctgttcggaacatgabatcttggaagagtttcagact	1968
Qy	781	gtctvgsgcgcgcacatgcgaacagcatgcgtcttctgctctgctgtctgtgcacacctctgctga	840
Db	1969	gtctvgsgcgcgcacatgcgaacagcatgcgtcttctgctctgctgtgtgcacacctctgctga	2028
Qy	841	aagrgtgtctatctccgggggaacacatgcccctgcgagsgctctgtctccggatgagggaagagt	900
Db	2029	aagrgtgtctatctccgggggaacacatgcccctgcgagsgctctgtctccggatgagggaagagt	2088
Qy	901	ccacccctctctgatacatagaaagccacacctgtgaagatgttcttggaagaagagagcagtgcga	960
Db	2089	ccacccctctctgatacatagaaagccacacctgtgaagatgttcttggaagaagagagcagtgcga	2148
Qy	961	agacacacagacacaaagctcccaagccatcagctgagtgagt	1001
Db	2149	agacacacagacacaaagctcccaagccatcagctgagtgagtgagt	2189

RESULT	3
AA52810	standard; cDNA; 2478 BP.
AC	AAA52810;
AD	20-SEP-2000 (first entry)
AE	Human sulphatase G cDNA.
AF	Human sulphatase G; hSG; chromosome 17p11.2; gene therapy; ss.
AG	Homo sapiens.
AH	Key
AI	Location/Qualifiers
AK	CDS
AL	1..2478
AM	/*tag= a
AN	/partial
AO	/product= "hSG"
AP	MO200034327-A1.
AQ	15-JUN-2000.
AR	09-DEC-1999; 99MO-A001092.
AS	09-DEC-1998; 98AU-0007624.
AT	(WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.
AW	Hopwood JJ, Litjens T, Hu RL;
AX	WPI; 2000-431273/37.
AY	P-PSDB; AAY99850.
BA	Novel isolated DNA sequence which encodes human sulfatase G or its
BB	fragment useful in gene therapy for treating patients suffering from
BC	sulfatase deficiency -
BD	Claim 2; Page 29-30; 33pp; English.
BE	The present sequence encodes human sulphatase G (hSG). hSG is
BF	not a member of the well-characterised CTRP sulphatase family.
BG	It belongs to a family showing sequence similarity to a sulphatase
BH	from the marine bacterium Alteromonas carrageneovora. The hsg gene
BI	contains 23 exons and is located at chromosome 17p11.2. The present
BJ	sequence is clone lambda29.1 from a human testes cDNA library. It was
BK	isolated using human EST sequences with sequence similarity to the
BL	non-CTPSP family as a probe to screen the library. The cDNA insert was
BM	subcloned and the DNA sequence of both strands was determined. The
BN	sequence may be used to treat a patient suffering from hSG deficiency

CC by replacing, repairing, or compensating for a DNA sequence within that
 CC patient's genome.
 XX
 SQ Sequence 2478 BP; 587 A; 686 C; 709 G; 496 T; 0 other;

Query Match 98.3%; Score 984.2; DB 21; Length 2478;
 Best Local Similarity 99.6%; Pred. No. 2,4e-265;

Matches 997; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 tccagcccaagaattcaaacccagctcaacatccatccaccgagacatcttccccctgct 60
 |||||||
 Db 1140 tccagcccaagaattcaaacccagctcaacatccatccaccgagacatcttccccctgct 1199
 QY 61 caccagttccgctgttaagaaggagggcccaacctcagttgcccattgttcagggtga 120
 |||||||
 Db 1200 caccagttccgctgttaagaaggagggcccaacctcagttgcccattgttcagggtga 1259
 QY 121 atgctctcctaagtaaccagctccgcccagaaggaggtgcagaaggatgcattattac 180
 |||||||
 Db 1260 atgctctcctaagtaaccagctccgcccagaaggaggtgcagaaggatgcattattac 1319
 QY 181 ttgcaatccctgaggaattcattgttagtgagcgctgcagctcccaactccagccagagcgt 240
 |||||||
 Db 1320 ttgcaatccctgaggaattcattgttagtgagcgctgcagctcccaactccagccagagcgt 1379
 QY 241 gcagaggtacaggaaggatgcagagagcgccagcccaagcagagaaaaagatcagta 300
 |||||||
 Db 1380 gcagaggtacaggaaggatgcagagagcgccagcccaagcagagaaaaagatcagta 1439
 QY 301 cccagaaatcatcttctctgtgaacaaggctgcacatcccgatgaagaattcgaaatgtcag 360
 |||||||
 Db 1440 cccagaaatcatcttctctgtgaacaaggctgcacatcccgatgaagaattcgaaatgtcag 1499
 QY 361 tgcacaatctgtcaacaagaagcccgagcagctctctgtcactgtgacgtgtgaggcac 420
 |||||||
 Db 1500 tgcacaatctgtcaacaagaagcccgagcagctctctgtcactgtgacgtgtgaggcac 1559
 QY 421 atttggcagcgtgtgccgcatattacgagacagagtggaagggtccctgggcaaccctggc 480
 |||||||
 Db 1560 gtttggcagcgtgtgccgcatattacgagacagagtggaagggtccctgggcaaccctggc 1619
 QY 481 tgcctgtgtgtgtcccaactgtgcagcagcagataccaacacgggcttgcagaattctgc 540
 |||||||
 Db 1620 tgcctgtgtgtgtcccaactgtgcagcagcagataccaacacgggcttgcagaattctgc 1678
 QY 541 tgcagagagaagcgcgctgtgcacatcttgggaaggcgcttcacaccttgcgtgtgtg 600
 |||||||
 Db 1679 tgcagagagaagcgcgctgtgcacatcttgggaaggcgcttcacaccttgcgtgtgtg 1738
 QY 601 cccccaacacagctcaaaagcctgtgctccagcagtaaccaacagtgccaagagttcctgc 660
 |||||||
 Db 1739 cccccaacacagctcaaaagcctgtgctccagcagtaaccaacagtgccaagagttcctgc 1798
 QY 661 accaatcagatgatgtctctgcacaatgccttaaggaaagggtctgagatccagttcctg 720
 |||||||
 Db 1799 accaatcagatgatgtctctgcacaatgccttaaggaaagggtctgagatccagttcctg 1858
 QY 721 cagtggaagatgatcagttgctgtgttggaacatgtgatttgggaagatttcagacct 780
 |||||||
 Db 1859 cagtggaagatgatcagttgctgtgttggaacatgtgatttgggaagatttcagacct 1918
 QY 781 gttctgtgtgcacgtgcaagcagatgctttggcgtgtgcgtgcacacacctgcgtctgga 840
 |||||||
 Db 1919 gttctgtgtgcacgtgcaagcagatgctttggcgtgtgcgtgcacacacctgcgtctgga 1978
 QY 841 aagttgtctatttcgggggaacacatgccttcgagagctctgtgtccgatgtgggaaagatg 900
 |||||||
 Db 1979 aagttgtctatttcgggggaacacatgccttcgagagctctgtgtccgatgtgggaaagatg 2038
 QY 901 ccaacctctgtatactagaagcacccttggaagatgttcttggaaagggaagcagttgaaa 960
 |||||||
 Db 2039 ccaacctctgtatactagaagcacccttggaagatgttcttggaaagggaagcagttgaaa 2098

QY 961 agacacacagcacaacgtlcccaagccatccagcgttggggatg 1001
 |||||||
 Db 2099 agacacacagcacaacgtlcccaagccatccagcgttggggatg 2139

RESULT 4
 AAA60207
 ID AAA60207 standard; DNA; 2664 BP.
 XX
 AC AAA60207;
 XX
 DT 07-DEC-2000 (first entry)
 XX
 DE Human prostate cancer predisposing gene HPC2 genomic sequence.
 XX
 KW Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
 KW gene therapy; peptide therapy; drug design; ds.
 OS Homo sapiens.
 OS
 FH Key
 FT CDS
 FT
 FT Location/Qualifiers
 FT 910..26039
 FT /tag= a
 FT /product= "HPC2"
 FT /note= "this sequence contains introns"
 FT /transl_except= (pos:23892..23895,aa:Glu)
 FT 910..1154
 FT /tag= b
 FT /number= 1
 FT 1736..1786
 FT /tag= c
 FT /number= 2
 FT 1925..1995
 FT /tag= d
 FT /number= 3
 FT 3025..3089
 FT /tag= e
 FT /number= 4
 FT 4361..4418
 FT /tag= f
 FT /number= 5
 FT 5582..5650
 FT /tag= g
 FT /number= 6
 FT 7075..7194
 FT /tag= h
 FT /number= 7
 FT 8186..8244
 FT /tag= i
 FT /number= 8
 FT 12878..12936
 FT /tag= j
 FT /number= 9
 FT 13032..13104
 FT /tag= k
 FT /number= 10
 FT 13756..13868
 FT /tag= l
 FT /number= 11
 FT 15283..15378
 FT /tag= m
 FT /number= 12
 FT 16278..16416
 FT /tag= n
 FT /number= 13
 FT 16498..16583
 FT /tag= o
 FT /number= 14
 FT 18583..18701
 FT /tag= p
 FT /number= 15
 FT 20349..20445
 FT exon

```
FT      /*tag= q  
FT      /number= 16  
FT      22172..22310  
FT      exon  
FT      /*tag= r  
FT      /number= 17  
FT      22879..22917  
FT      exon  
FT      /*tag= s  
FT      /number= 18  
FT      23045..23154  
FT      exon  
FT      /*tag= t  
FT      /number= 19  
FT      23795..23895  
FT      exon  
FT      /*tag= u  
FT      /number= 20  
FT      23973..24093  
FT      exon  
FT      /*tag= v  
FT      /number= 21  
FT      24354..24432  
FT      exon  
FT      /*tag= w  
FT      /number= 22  
FT      25026..25170  
FT      exon  
FT      /*tag= x  
FT      /number= 23  
FT      25812..26036  
FT      exon  
FT      /*tag= y  
FT      /number= 24  
FT      26447..26452  
FT      polyA_signal  
FT      /*tag= z  
XX      WO200027864-A1.  
XX      18-MAY-2000.  
XX      PD  
XX      05-NOV-1999; 99WO-US26055.  
XX      PE  
XX      06-NOV-1998; 98US-0107468.  
XX      PR  
XX      (MYRI-) MYRIAD GENETICS INC.  
XX      PI  
XX      Tavligian SV, Teng DHF, Simard J, Rommens JM;  
XX      DR  
XX      WPI: 2000-376481/32.  
XX      P-PSDB; AAB07228.  
XX      PT  
XX      Human prostate cancer (HPC)2 nucleic acids, polypeptides, and  
XX      antibodies, useful for treatment and diagnosis of prostate cancer  
XX      Claim 3; Page 108-122; 157pp; English.  
XX      PS  
XX      The present sequence is the genomic sequence of the human prostate  
XX      cancer predisposing gene HPC2, which is found on chromosome 17p. Some  
XX      alleles of this gene cause a predisposition to cancer, particularly  
XX      prostate cancer. This gene and its protein can be used in peptide and  
XX      gene therapy for cancer patients, as well as being useful as diagnostic  
XX      tools (both for cancer sufferers and those with a predisposition to the  
XX      disease) and in the production of cancer drugs. This sequence was  
XX      isolated by cloning and sequencing the region of the genome which  
XX      appeared to cause a predisposition to prostate cancer.  
XX      CC  
XX      Sequence 26664 BP; 6173 A; 6300 C; 6519 G; 7661 T; 11 other;  
XX      SO  
Query Match 13.6%; Score 136.6; DB 21; Length 26664;  
Best Local Similarity 73.9%; Pred. NO. 6.2e-28;  
Matches 224; Conservative 1; Mismatches 0; Indels 78; Gaps 1;  
QY 668 cagatgattcttcgccaatgccttcagaagggcgctagatctcagtcctgcagtgga 727  
Db 23792 cagatgattcttcgccaatgccttcagaagggcgctagatctcagtcctgcagtgga 23851  
QY 728 aagattgacagttcgtctgtgcgaacatgtgattggaag----- 768  
Db 23852 aagattgacagttcgtctgtgcgaacatgtgattggaaggtgaagggcgacagccgc 23911
```

```
QY 769 -----a 769  
Db 23912 aggcacatgaggcgaggtgggagcagagctgcagagccctccagcccaacctttaa 23971  
QY 770 gttcagacctgtctgtgcgagcactgcgaacatgcgttggctgtgcgtgtgcacac 829  
Db 23972 gttcagacctgtctgtgcgagcactgcgaacatgcgttggctgtgcgtgtgcacac 24031  
QY 830 ctctgctggaagtggtctatctccggggacacacatgccttcgagagctgtgtccgat 889  
Db 24032 ctctgctggaagtggtctatctccggggacacacatgccttcgagagctgtgtccgat 24091  
QY 890 ggg 892  
Db 24092 ggg 24094  
RESULT 5  
AAC34173  
ID AAC34173 standard; DNA; 1617 BP.  
AC AAC34173;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 5716.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW Protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.
```

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145219.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152263.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155186.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.


```
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147410.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151436.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
```

PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 54.6%; Score 41.6; DB 21; Length 464;
Matches 83; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

OY 281 gcgagaaagagtcagtcacccagaatcatcttcttggaacagggttcgcatccg 340
DB 89 gaaagattatcacgagatgcacgtgagatgttattcttaaggcacgtgtctgcagcct 148
OY 341 atgaagattcgaatgtcagtcacacattgtcaacataagcccgacagctctgcta 400
DB 149 tcaaataccgaaacgttagtgcaattttattgactgttctctcagaggtagttcttc 208
OY 401 ctggaactgtgtgagggcacatttgagcagct 432
DB 209 ctgattgtgtgaggaaccttgagccagct 240

RESULT 8

AAC04378
ID AAC04378 standard; cDNA; 521 BP.

AC AAC04378;
XX
DT 06-OCT-2000 (first entry)
XX

DE Human secreted protein 5' EST, SEQ ID NO: 8453.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1: SEQ ID 8453; 71bp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

XX Sequence 521 BP; 122 A; 187 C; 87 G; 123 T; 2 other;

Query Match
Best Local Similarity 3.8%; Score 37.6; DB 21; Length 521;
Matches 67; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 30 acctcatccaccggacatcttccctcgtccaccagttccgtgtaagaagagggcc 89
DB 187 acctatcacctggaccgctgtgctctcctcaaccttttggccagagattgttccacc 246
OY 90 ccaacctcagatgtgcccattggtcagggtgaatgctccctcaagtaccagctccgt 145
DB 247 acctctcctccacccttatgttccaggagattccaccctcctcctccgacccct 302

RESULT 9

AAC89621
ID AAC89621 standard; DNA; 2517 BP.

AC AAC89621;

XX
DT 08-MAR-2001 (first entry)
XX

DE S. cerevisiae YKR079C gene.

KW Yeast; germination; proliferation; essential gene; antifungal agent;
KW insecticide; herbicide; anti-proliferation drug; cancer; psoriasis;
KW restenosis; YKR081C; YFR003C; YGR277C; YGR278W; YKR071C; YKR079C;
KW YKR083C; ds.

OS Saccharomyces cerevisiae.

PN W0200071161-A1.

PD 30-NOV-2000.

PF 12-MAY-2000; 2000WO-US13017.

PR 21-MAY-1999; 99US-0315794.

PR 02-SEP-1999; 99US-0389341.

XX (ROSE-) ROSETTA INPHARMATICS INC.

XX Roberts CJ;

XX WPI: 2001-025092/03.
XX P-PSDB; AAB49964.

PT Identifying antifungal compounds which target yeast essential genes
PT comprises use of novel Saccharomyces cerevisiae essential genes
PT YFR003C, YGR277C, YGR278W, YKR0701C, YKR079C or YKR083C or YKR081C -
XX
XX Example 5; Fig 29; 127pp; English.

CC The present invention provides methods of identifying antifungal agents
CC using the coding and protein sequences of several yeast genes. These are
CC essential for the germination and proliferation of Saccharomyces
CC cerevisiae, and include YKR081C, YFR003C, YGR277C, YGR278W, YKR071C,
CC YKR079C and YKR083C. The sequences can also be used to identify compounds
CC for use as herbicides, insecticides and anti-proliferation drugs which
CC can be used in the treatment of cancer, psoriasis and restenosis. This is
CC because they can be used to identify plant, insect and human homologues
CC of the yeast genes.

XX Sequence 2517 BP; 881 A; 417 C; 496 G; 723 T; 0 other;

Query Match
Best Local Similarity 3.7%; Score 37; DB 22; Length 2517;
Matches 67; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 883 tccgagatgggaaagatgccacctctctgatacatgaagccacctggaagatggttg 942
DB 2126 tcgaataagctataattcagatcatatttaattcaagagactacataatcagctac 2185

PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130049.
PR 23-APR-1999; 990S-01300510.
PR 28-APR-1999; 990S-01300891.
PR 30-APR-1999; 990S-0131049.
PR 04-APR-1999; 990S-0132048.
PR 05-MAY-1999; 990S-0132407.
PR 06-MAY-1999; 990S-0132484.
PR 06-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132487.
PR 11-MAY-1999; 990S-0132863.
PR 14-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.

PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 21-JUL-1999; 990S-0144684.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 22-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145513.
PR 27-JUL-1999; 990S-0145818.
PR 27-JUL-1999; 990S-0145919.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 06-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150586.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155569.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.

PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160766.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	3.6%;	Score 36;	DB 21;	Length 1409;
Best Local Similarity	60.0%;	Pred. NO. 2.5;		
Matches 60;	Conservative 0;	Mismatches 40;	Indels 0;	Gaps 0;

Qy 574 agccgcctcaaccccttgcgtggttcccccaaccagctcaaaagcctgtgcctcaagcagt 6333

Db 350 AGAAGCTGAACCACTGATGATTTGACCCCTGTACCTGCACCGAAGCCGCTTCCAGTAGA 2911

Qy	634	accacaaccagctgccaagagctcctgtcaaccacatcagtat	673
Db	290	acctgaaccagatgacccctgatctctgcacccaacccgatt	251

RESULT	12
AAA59146/c	
ID	AAA59146 standard; DNA; 19053 BP.

AC AAA59146;

DT 07-NOV-2000 (first entry)

DE Nucleotide sequence of the antibiotic Tet-aviv gene cluster.

KW polyketide; antibiotic Tel-Aviv; cell wall synthesis;

lipid-disaccharide-pentapeptide; gingivitis; ss

05 *Myxococcus xanthus*.

FH	Key	Location/Qualifiers
FT	CDS	complement (2955..601)

FT	FT	CDS
	/note= "encodes AAB07666"	
	complement (4702..3116)	

FT	FT	FT	CDS
		complement (5595..4720)	/note= "encodes AAB07667"

FT	FT	FT	CDS
6290..6793	/note= "encodes AAB07666"		

	FT	FT	FT	CDS
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9
10	10	10	10	10
11	11	11	11	11
12	12	12	12	12
13	13	13	13	13
14	14	14	14	14
15	15	15	15	15
16	16	16	16	16
17	17	17	17	17
18	18	18	18	18
19	19	19	19	19
20	20	20	20	20
21	21	21	21	21
22	22	22	22	22
23	23	23	23	23
24	24	24	24	24
25	25	25	25	25
26	26	26	26	26
27	27	27	27	27
28	28	28	28	28
29	29	29	29	29
30	30	30	30	30
31	31	31	31	31
32	32	32	32	32
33	33	33	33	33
34	34	34	34	34
35	35	35	35	35
36	36	36	36	36
37	37	37	37	37
38	38	38	38	38
39	39	39	39	39
40	40	40	40	40
41	41	41	41	41
42	42	42	42	42
43	43	43	43	43
44	44	44	44	44
45	45	45	45	45
46	46	46	46	46
47	47	47	47	47
48	48	48	48	48
49	49	49	49	49
50	50	50	50	50
51	51	51	51	51
52	52	52	52	52
53	53	53	53	53
54	54	54	54	54
55	55	55	55	55
56	56	56	56	56
57	57	57	57	57
58	58	58	58	58
59	59	59	59	59
60	60	60	60	60
61	61	61	61	61
62	62	62	62	62
63	63	63	63	63
64	64	64	64	64
65	65	65	65	65
66	66	66	66	66
67	67	67	67	67
68	68	68	68	68
69	69	69	69	69
70	70	70	70	70
71	71	71	71	71
72	72	72	72	72
73	73	73	73	73
74	74	74	74	74
75	75	75	75	75
76	76	76	76	76
77	77	77	77	77
78	78	78	78	78
79	79	79	79	79
80	80	80	80	80
81	81	81	81	81
82	82	82	82	82
83	83	83	83	83
84	84	84	84	84
85	85	85	85	85
86	86	86	86	86
87	87	87	87	87

FT	/tag- e
FT	/note= "encodes AAB07670"
FT	719..8378
CDS	

FI	/tag= I
FT	/note= "encodes AAB07671"
FT	8404 9378
CDS	

FT CDS

FT	/*tag= g
FT	/note="encodes AAB07672"
FT	9386..9364
FT	/tag= h
FT	/note="encodes AAB07673"
FT	9757..10878
FT	/tag= i
FT	/note="encodes AAB07674"
FT	10909..11421
FT	/tag= j
FT	/note="encodes AAB07675"
FT	11473..12897
FT	/tag= k
FT	/note="encodes AAB07676"
FT	12938..13891
FT	/tag= l
FT	/note="encodes AAB07677"
FT	13809..14898
FT	/tag= m
FT	/note="encodes AAB07678"
FT	14963..16213
FT	/tag= n
FT	/transl_except= (pos: 15521..15523, aa: Lys)
FT	/note="encodes AAB07679"
FT	16224..17009
FT	/tag= o
FT	/transl_except= (pos: 16593..16595, aa: Ser)
FT	/note="encodes AAB07680"
FT	17000..17767
FT	/tag= p
FT	/note="encodes AAB07681"
FT	17782..19053
FT	/tag= q
FT	/note="encodes AAB07682"
PN	EP1026248-A2.
PD	09-AUG-2000.
XX	
XX	31-JAN-2000; 2000EP-0300747.
PF	
PR	29-JAN-1999; 99US-0240537.
XX	
PA	(UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX	
PI	Rosenberg E, Ron E, Orr E, Paltan Y;
DX	
DR	WPI: 2000-500254/45.
DR	P-PSDB; AAB07666, AAB07667, AAB07668, AAB07669, AAB07670, AAB07671,
DR	AAB07672, AAB07673, AAB07674, AAB07675, AAB07676, AAB07677, AAB07678,
DR	AAB07679, AAB07680, AAB07681, AAB07682.
XX	
PT	Novel DNA sequence involved in polyketide antibiotic Tel-Aviv
PT	production useful for inhibiting cell wall synthesis and in wide range
PT	of clinical applications such as treating gingivitis -
XX	
PS	Disclosure; Page 12-19; 66pp: English.
XX	
CC	The specification describes a DNA sequence which partially encodes
CC	a functional portion of polypeptide component required for synthesizing
CC	the polyketide antibiotic Tel-Aviv, postmodification of antibiotic
CC	Tel-Aviv, or regulation of biosynthesis of antibiotic Tel-Aviv. The
CC	antibiotic Tel-Aviv is a macrocyclic polyketide synthesised through
CC	the incorporation of acetate, methionine, and glycine. It inhibits cell
CC	wall synthesis by interfering with the polymerisation of the
CC	lipid-disaccharide-pentapeptide. Antibiotic Tel-Aviv genes are useful
CC	in combinatorial genetics, and for encoding protein components for the
CC	synthesis, modification and regulation of antibiotic antibiotic Tel-Aviv
CC	Antibiotic Tel-Aviv is useful in a wide range of clinical applications
CC	such as treating gingivitis. Antibiotic Tel-Aviv is also useful for
CC	generating new biological agents from its secondary metabolites. The
CC	present sequence encodes proteins involved in synthesis of antibiotic
CC	Tel-Aviv.

CC and (7) transgenic plants that contain cells of (2). (1), and their
CC fragments, are useful for expression of recombinant lipopeptide
CC synthases, and as probes and primers for detection, isolation and
CC amplification of full-length cDNA sequences. (1) are used to produce
CC transgenic plants. Lipopeptide synthases are useful for production of
CC antiviral, antibacterial or antifungal lipopeptides, particularly useful
CC for treatment of plants, and for production of lipopeptide surfactants
CC used for increasing bioavailability in microbial depollution processes
CC (acting as emulsifiers) and in tertiary crude oil recovery. Individual
CC subunits from different lipopeptide synthases may be combined to allow
CC synthesis of many different biologically active substances.

Query Match	3.5%	Score 35.2	DB 22	Length 1716
Best Local Similarity	49.0%	Pred. No. 4.5		
Matches 94; Conservative	0	Mismatches 98	Indels 0	Gaps 0

QY	30	acccatcaaccgcggaaatcttccctctgtacacagttccgcgtgtaaagagagcc	89
	11	acccatgatacccttccacacttcgtcggacaaatgvgcgtgaataccgagcgaagcc <th>70</th>	70
QY	90	ccacccctcagtgctgccatggttcagggtaatgacctccctccaaagtacacagctccgtccca	149
Db	71	tggagacctatcgtgcgtccgcgaagagatcggagagcgtccacagtcgtcttcttctc	130
QY	150	ggaaggaagtggcagaagagatgcattatcttgcacatcccttgaggaattcatgttagg	209
Db	131	ggggcatctggccagctcgcttggcgacaatgctcgagatgctgagcatcatattgattgtg	190
QY	210	cgctgcagcgttc	221
Db	191	aaatgcagagttc	202

```

RESULT 15
AAQ39018
ID      AAQ39018 standard; DNA; 3474 BP

```

AAQ39018;

DT 13-AUG-1993 (first entry)

DE Sequence of pure maize optimized synthetic Bt CRYIA(b) gene.

KW Bacillus thuringiensis; CRYIA(b); insecticide; ss

05 Synthetic.

PN WO9307278-A.

PD 15-APR-1993.

PF 05-OCT-1992; 92WO-US08476.

PR 04-OCT-1991; 91US-0772027.

XX

XX

PI Launiskl, Lewis KS, Merlín EJ, Rothstein SJ, Warren GW;

XX
3

XX
XX

PT DNA sequences, enclosing insecticidal proteins for controlling
PT Coleoptera and Lepidoptera pests - optimised for expression in
PT plants, esp. maize

PS Disclosure; Pages 167-169; 286pp; English

CC The inventors claim a maize optimised coding sequence for *Bacillus*
CC *thuringiensis* Bt) insecticidal protein CryIa(b). Claimed sequences
CC are AAQ39019 and AAQ39020.

SQ Sequence 3474 BP; 749 A; 1255 C; 1023 G; 447 T; 0 other;

Query Match	3.5%	Score 35	DB 14	Length 3474
Best Local Similarity	46.2%	Pred. No.	6.8	
Matches 116	Conservative	0	Mismatches 135	Indels 0
			Gaps	0

QY	27	tcaacccatcacaccccgagatattcccccctgtacacccagtttccgcgtgtgaagagagag	86
Db	2468	tcaagatacaagaaccacagagacgaccacgcccgcctgtgtcaccccgagttctctctgtgagaga	252
QY	87	gccccacaccttaagtgltgcccatgtgtctcaaggtgtgaatgacctctcagaatcacacgtcttc	146
Db	2528	agccccctgtgtgtggtgagagccctctgtcccgctgtgaagtcggtcgagagaaggtgtgtcgcgaga	258
QY	147	ccaaagagtgagtggtgcagaagagatgccattcttctgtcaatccctgtgagagattcatagttg	206
Db	2568	agcgtgagaaagctgtgaggtgtgagaaccaacatctgtgatacaagaagagccaaagagagctgtg	264
QY	207	aggtgtcttcagacttccccaacttccacgacagagctgtgcacagatcacaggaagaggtgtgcag	266
Db	2648	agcgcctgtgtgtgaataacgcagtcagaccgcctgtgaggtcgcaacacacatctgcata	270
QY	267	acgtgcccacgtc	27
Db	2708	tccacgtcgtcc	2718

Search completed: October 28, 2001, 21:49:58
Job time: 2731 sec

; SOFTWARE: PatentIn Ver. 2.0


```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,448A
CLASSIFICATION: 800
FILING DATE: 02-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/951,715
FILING DATE: 25-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40403
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8582
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 3546 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3543
OTHER INFORMATION: /product= "Full-length, hybrid,
OTHER INFORMATION: maize optimized heat stable cryIA(b)"
OTHER INFORMATION: /note= "Disclosed in Figure 13 as contained in pcib5513."
US-08-459-448A-14

```

```

Query Match 3.5%; Score 35; DB 2; Length 3546;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 116; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 27 tcacccatccaccgccgacatcttccccctgcgtaccagtttcgctgtaagaaggagg 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2546 TCAAGATCAACACCCAGACGCCACGCCGCCCTGCGCAATCTAGAGTTCTCGAGAGAGA 2605

QY 87 gcccaccctcagtggtgccatggttcagggtgaatgctctcctcaagtaaccagctcgtc 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2606 AGCCCTCTGTGGGGAGAGCCCTGGGCCCGGGAAGCGCGGAGAAAGAGTGGCGGACA 2665

QY 147 ccaggaggagtgagcagagagatgccattatctgaatcctcgaaggaattcatagttg 206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2666 AGCCGAGAAAGCTGAGTGGAGACCAACATCGTGTACAGAGAGGCCCAAGAGAGAGCTGG 2725

QY 207 aggcgcctcagctccaccctccagcagcagcgfbcagagtgtaacaggagggtgcagg 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2726 AGGCCCTCTGTGTGAACAGCCAGTACGACGCCCTGACGCGACCAACATCGGCATGA 2785

QY 267 acggcccgagcc 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2786 TCCACGCCGCC 2796

```

```

RESULT 5
US-08-459-595A-14
; Sequence 14, Application US/08459595A
; Patent No. 6018104
; GENERAL INFORMATION:
; APPLICANT: Kozziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.

```

```

APPLICANT: Warren, Gregory W.
APPLICANT: Evola, Stephen V.
APPLICANT: Crossland, Lyle D.
APPLICANT: Wright, Martha S.
APPLICANT: Merlin, Ellis J.
APPLICANT: Launis, Karen L.
APPLICANT: Rothstein, Steven J.
APPLICANT: Bowman, Cindy G.
APPLICANT: Dawson, John L.
APPLICANT: Dunder, Erik M.
APPLICANT: Pace, Gary M.
APPLICANT: Suttle, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6018104artis Corporation
STREET: Patent & Trademark Dept., 520 White Plains
STREET: Rd., POB 2005
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-9005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,595A
FILING DATE: 02-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/951,715
FILING DATE: 25-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40403
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8582
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 3546 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3543
OTHER INFORMATION: /product= "Full-length, hybrid,
OTHER INFORMATION: maize optimized heat stable cryIA(b)"
OTHER INFORMATION: /note= "Disclosed in Figure 13 as contained in pcib5513."
US-08-459-595A-14

```

```

Query Match 3.5%; Score 35; DB 3; Length 3546;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 116; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 27 tcacccatccaccgccgacatcttccccctgcgtaccagtttcgctgtaagaaggagg 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2546 TCAAGATCAACACCCAGACGCCACGCCGCCCTGCGCAATCTAGAGTTCTCGAGAGAGA 2605

QY 87 gcccaccctcagtggtgccatggttcagggtgaatgctctcctcaagtaaccagctcgtc 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 2606 AGCCCCGTGTGGCGAGGCCCCGTGGCCGTGAGAGCCGCCGAGAGAAAGTGGCGGACCA 2665
QY 147 ccagagagagagtgccagagagatgcatcttacttgcgaatccttgaggaattcatagttg 206
Db 2666 AGCGCGAAGAGCTGGAGTGGAGACCAACATCTGTACAGAGAGGCCAAGAGAGCTGG 2725
QY 207 aggcgtgcagcttcccaacttccagcagagcgctgcagaggtacaggaagtgcgcaag 266
Db 2726 AGCGCCTGTTCTGAGACAGCAGTACGACCGCTCGAGAGGCCGACACCAACATCGCCATGA 2785
QY 267 acgagcccaagcc 277
Db 2786 TCCACGCCCGCC 2796

RESULT 6
US-08-459-504B-14
; Sequence 14, Application US/08459504B
; Patent No. 6075185
; GENERAL INFORMATION:
; APPLICANT: Kozziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttle, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 6075185artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,504B
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/459,595
; FILING DATE: 02-JUN-1995
; APPLICATION NUMBER: US 07/951,715
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1577/CJP/DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8587
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3546 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3543
OTHER INFORMATION: /product= "Full-length, hybrid,
OTHER INFORMATION: maize optimized heat stable cry1A(b)"
OTHER INFORMATION: /note= "Disclosed in Figure 13 as contained in pCIB513."
US-08-459-504B-14

Query Match 3.5%; Score 35; DB 3; Length 3546;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 116; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 27 tcaacctcattccaccggaatcttccctctgctcaacagttccgctgtaagaagag 86
Db 2546 TCAGATCAGAGCCAGACGAGCGCCACGCCCTGGGCAATCTAGACTTCTGTGAGAGAGA 2605
QY 87 gcccaacctcagtggtcccatggttcagggtgaatgctctcctcaagtaccagctccgtc 146
Db 2606 AGCCCTGTGTGGCGAGGCCCCGTGGCCGTGAGAGCGCCGAGAGAAAGTGGCGGACCA 2665
QY 147 ccagagagagagtgccagagagatgcatcttacttgcgaatccttgaggaattcatagttg 206
Db 2666 AGCGCGAAGAGCTGGAGTGGAGACCAACATCTGTACAGAGAGGCCAAGAGAGCTGG 2725
QY 207 aggcgtgcagcttcccaacttccagcagagcgctgcagaggtacaggaagtgcgcaag 266
Db 2726 AGCGCCTGTTCTGAGACAGCAGTACGACCGCTCGAGAGGCCGACACCAACATCGCCATGA 2785
QY 267 acgagcccaagcc 277
Db 2786 TCCACGCCCGCC 2796

RESULT 7
US-08-459-444-14
; Sequence 14, Application US/08459444A
; Patent No. 6121014
; GENERAL INFORMATION:
; APPLICANT: Kozziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED
; NUCLEIC ACID CODING SEQUENCE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 6121014artis Agribusiness Biotechnology Research, Inc.
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,444A
; FILING DATE: 02-JUN-1995

```

1 CLASSIFICATION: <unknown>
2
3 PRIOR APPLICATION DATA:
4
5 APPLICATION NUMBER: US 07/951,715
6 FILING DATE: 25-SEP-1992
7 APPLICATION NUMBER: US 07/772,027
8 FILING DATE: 04-OCT-1991
9
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Meigs, J. Timothy
12 REGISTRATION NUMBER: 38,241
13 REFERENCE/DOCKET NUMBER: S-18805/P1/CGC1577/CIP/DIV6
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: (919)541-8587
16 TELEFAX: (919)541-8689
17
18 INFORMATION FOR SEQ ID NO: 14:
19
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 3546 base pairs
22 TYPE: nucleic acid
23 STRANDEDNESS: single
24 TOPOLOGY: linear
25
26 MOLECULE TYPE: other nucleic acid
27 DESCRIPTION: /desc = "synthetic DNA"
28
29 HYPOTHETICAL: NO
30
31 FEATURE:
32
33 NAME/KEY: CDS
34 LOCATION: 1..3543
35
36 OTHER INFORMATION: /product= "Full-length, hybrid,
37 malte optimized heat stable cyaI(b)"
38
39 /note= "disclosed in Figure 13 as contained in pcib5513."
40
41 SEQUENCE DESCRIPTION: SEQ ID NO: 14:
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
```

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/053,549
 FILING DATE: 01-APR-1998
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Pace, Gary M.
 REGISTRATION NUMBER: 40,403
 REFERENCE/DOCKET NUMBER: CGC 1995
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8582
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8854 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 11..3694
 OTHER INFORMATION: /product= "hyfLib protein"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 8854
 OTHER INFORMATION: /note= "Sequence of PCIB5520
 Patent No. 6121521
 OTHER INFORMATION: containing coding sequence for hyfLib protein"
 US-09-053-549-1

```

? APPLICANT: Wright, Martha S.
? APPLICANT: Merlin, Ellis J.
? APPLICANT: Launis, Karen L.
? APPLICANT: Rothstein, Steven J.
? APPLICANT: Bowman, Cindy G.
? APPLICANT: Dawson, John L.
? APPLICANT: Dunder, Erik M.
? APPLICANT: Pace, Gary M.
? APPLICANT: Suttie, Janet L.
? TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
? TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
? NUMBER OF SEQUENCES: 94
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: CIBA-GEIGY Corporation
? STREET: 7 Skyline Drive
? CITY: Hawthorne
? STATE: New York
? COUNTRY: USA
? ZIP: 10532
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30B
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/07/951,715A
? FILING DATE: 25-SEP-1992
? CLASSIFICATION: 800
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/772,027
? FILING DATE: 04-OCT-1991
? ATTORNEY/AGENT INFORMATION:
? NAME: Spruill, W. Murray
? REGISTRATION NUMBER: 32,943
? REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (919)541-8615
? TELEFAX: (919)541-8689
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3468 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "Synthetic DNA"
? HYPOTHEICAL: NO
? FEATURE:
? NAME/KEY: misc.feature
? LOCATION: 1..3468
? OTHER INFORMATION: /product- "Full length pure maize
? OTHER INFORMATION: optimized synthetic bt"
? OTHER INFORMATION: /note- "Disclosed in Figure 3 as syn1T.mze"
US-07-951-715A-2

Query Match 3.3%; Score 33.4; DB 1; Length 3468;
Best Local Similarity 45.8%; Pred. No. 3.1;
Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 27 tcaactcatcacccgagacattccctgtctcacacagttcgcgtgaagaagaag 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2468 TCAAGATCAAGACCCAGCGGCGCGCTGGGCACTGAGATTCTCTGGAGAGA 2527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 87 gcccaacctcagtgtgcccattgttcagggtgaatgctctccaagtaccagtcctc 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2528 AGCCCTGTGGGGCGGAGCGCTGTGCGCGCTGAAGCGCGCGGAGAGAAAGTGGCGCACA 2587
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 147 ccagagagagatgcagagatgcattatcttgcacatcctcgaagaaattcattgt 206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2588 AGCGCAGAAAGCTGGAGTGGAGACCAACACTCTGTACAAAGAGAGCCCAAGAGAGCTGG 2647
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy- 207 aggcgtgcagcttcccaacttccagcagagcgtgcagaggtacagagagagtgcgcaag 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Db 2648 ACCCCCTGTCTGTAAGACGACGACGCGCTGCGCGCGGACACCAACATCGCATGA 2707
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 267 acggccacgac 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2708 TCCAGCGCGCC 2718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-07-951-715A-4
; Sequence 4, Application US/07951715A
; Patent No. 5625136
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalin M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/951,715A
; FILING DATE: 25-SEP-1992
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3468 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..3468
; OTHER INFORMATION: /product- "Full length synthetic
; OTHER INFORMATION: maize optimized"
; OTHER INFORMATION: /note- "Disclosed in Figure 3 as synful.mod. This sequenc
```

OTHER INFORMATION: identical to flsynbl.fin as disclosed in Figure 1."

US-07-951-715A-4

Query Match 3.3%; Score 33.4; DB 1; Length 3468;

Best Local Similarity 45.8%; Pred. No. 3.1;

Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 27 tcaacctatccaccgacatcttccccctgctcaccagtttcgcgtgaagaagg 86

Db 2468 TCAGATTAAGACCCAGACGCCACGCCCGCTGGCAACCTGGAGTTCTGTGAGAGA 2527

Qy 87 gccaccctcagtggtgcccagtggtcaggggtgaatgctctcctcagaccagctcc 146

Db 2528 AGCCCTGTGTGGGAGGAGCCCTGGCCCGGTGAACCGCGCGAGAGAAGTGGCCGACA 2587

Qy 147 ccaggaaggagtgagcagaggatgccattacttaactcactcaggaatcagattg 206

Db 2588 AGCCGAGAACGTGAGTGGAGACACCAACATGTTACAGAGAGGCCAAGAGAGCGTGG 2647

Qy 207 aggcgtcagctccaccctcagcagagcgtgcaagagtagcagagagtgctgcag 266

Db 2648 AGCCCTGTGTGGAACGCCAGTACGACCGCTGACGCGCGACACCAACATGCCATGA 2707

Qy 267 acggccagcc 277

Db 2708 TCCACGCCGCC 2718

RESULT 11

US-08-459-448A-2

Sequence 2, Application US/08459448A

Patent No. 5859336

GENERAL INFORMATION:

APPLICANT: Koziel, Michael G.

APPLICANT: Desai, Nalini M.

APPLICANT: Lewis, Kelly S.

APPLICANT: Kramer, Vance C.

APPLICANT: Warren, Gregory W.

APPLICANT: Eviola, Stephen V.

APPLICANT: Crossland, Lyle D.

APPLICANT: Wright, Martha S.

APPLICANT: Merlin, Ellis J.

APPLICANT: Launis, Karen L.

APPLICANT: Rothstein, Steven J.

APPLICANT: Bowman, Cindy G.

APPLICANT: Dawson, John L.

APPLICANT: Dunder, Erik M.

APPLICANT: Pace, Gary M.

APPLICANT: Suttie, Janet L.

TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESSES:

ADDRESS: Patent 6 Trademark Dept., 520 White Plains

STREET: Rd., POB 2005

CITY: Tarrytown

STATE: New York

COUNTRY: USA

ZIP: 10591-9005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/459,448A

FILING DATE: 02-JUN-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/951,715

FILING DATE: 25-SEP-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/772,027

FILING DATE: 04-OCT-1991

ATTORNEY/AGENT INFORMATION:

NAME: Pace, Gary M.

REGISTRATION NUMBER: 40403

REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIVA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919)541-8582

TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 3468 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "Synthetic DNA"

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: misc-feature

LOCATION: 1..3468

OTHER INFORMATION: /product= "Full-length pure maize

OTHER INFORMATION: optimized synthetic Bt"

OTHER INFORMATION: /note= "Disclosed in Figure 3 as syn1t.mze"

Query Match 3.3%; Score 33.4; DB 2; Length 3468;

Best Local Similarity 45.8%; Pred. No. 3.1;

Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 27 tcaacctatccaccgacatcttccccctgctcaccagtttcgcgtgaagaagg 86

Db 2468 TCAGATTAAGACCCAGACGCCACGCCCGCTGGCAACCTGGAGTTCTGTGAGAGA 2527

Qy 87 gccaccctcagtggtgcccagtggtcaggggtgaatgctctcctcagaccagctcc 146

Db 2528 AGCCCTGTGTGGGAGGAGCCCTGGCCCGGTGAACCGCGCGAGAGAAGTGGCCGACA 2587

Qy 147 ccaggaaggagtgagcagaggatgccattacttaactcactcaggaatcagattg 206

Db 2588 AGCCGAGAACGTGAGTGGAGACACCAACATGTTACAGAGAGGCCAAGAGAGCGTGG 2647

Qy 207 aggcgtcagctccaccctcagcagagcgtgcaagagtagcagagagtgctgcag 266

Db 2648 AGCCCTGTGTGGAACGCCAGTACGACCGCTGACGCGCGACACCAACATGCCATGA 2707

Qy 267 acggccagcc 277

Db 2708 TCCACGCCGCC 2718

RESULT 12

US-08-459-448A-4

Sequence 4, Application US/08459448A

Patent No. 5859336

GENERAL INFORMATION:

APPLICANT: Koziel, Michael G.

APPLICANT: Desai, Nalini M.

APPLICANT: Lewis, Kelly S.

APPLICANT: Kramer, Vance C.

APPLICANT: Warren, Gregory W.

APPLICANT: Eviola, Stephen V.

APPLICANT: Crossland, Lyle D.

APPLICANT: Wright, Martha S.

APPLICANT: Merlin, Ellis J.

APPLICANT: Launis, Karen L.

APPLICANT: Rothstein, Steven J.

APPLICANT: Bowman, Cindy G.

APPLICANT: Dawson, John L.

APPLICANT: Dunder, Erik M.

APPLICANT: Pace, Gary M.

APPLICANT: Suttle, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 585936artis Corporation
STREET: Patent & Trademark Dept., 520 White Plains
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-9005
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,448A
FILING DATE: 02-JUN-1995
CLASSIFICATION: 800
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/951,715
FILING DATE: 25-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40403
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8582
TELEFAX: (919)541-8689
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3468 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..3468
OTHER INFORMATION: /product= "Full length synthetic
OTHER INFORMATION: /note= "disclosed in Figure 3 as synful.mod. This sequence
OTHER INFORMATION: identical to flsynbt.fin as disclosed in Figure 1."
US-08-459-448A-4

Query Match 3.3%; Score 33.4; DB 2; Length 3468;
Best Local Similarity 45.8%; Pred. No. 3.1;
Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 27 tcaactcatccaccggaatcttccccctgtcaccagttccgtcgaagaagg 86
DB 2468 TCAAGATCAAGACCGAGGAGCGCCGCGCTGGGCACTGGAGTTCTTGAGAGGA 2527
QY 87 gcccacccctcagtggtccatagttcagggtgaatgcccctccatcacagctcgtc 146
DB 2528 AGCCCTGTGGTGGGAGAGCCCTGCGCTGTAGAGCCGCGGAGAAAGTGGCGCACA 2587
QY 147 ccaggagagagtgagcagagagatcattactgtcaatctcctgaggaattcatagtg 206
DB 2588 AGCGCGAAGAGCTGAGTGGGAGACCAACATCGTACAGAGGAGCCCAAGAGAGCGCTGG 2647
QY 207 aggcgctgcagctcccaacttcacagcagagcgtgcaggaatccaggagagtgcgacag 266
DB 2648 ACGCCCTGTTCGTGAACAGCCAGTACGACCGCTGCGAGGCCGACACCAACATGCGCATGA 2707

QY 267 acgcccagcc 277
DB 2708 TCACAGCCGCC 2718
RESULT 13
US-08-459-595A-2
Sequence 2, Application US/08459595A
Patent No. 6018104
GENERAL INFORMATION:
APPLICANT: Kozielec, Michael G.
APPLICANT: Desai, Nalin M.
APPLICANT: Lewis, Kelly S.
APPLICANT: Kramer, Vance C.
APPLICANT: Warren, Gregory W.
APPLICANT: Evola, Stephen V.
APPLICANT: Crossland, Lyle D.
APPLICANT: Wright, Martha S.
APPLICANT: Merlin, Ellis J.
APPLICANT: Launis, Karen L.
APPLICANT: Rothstein, Steven J.
APPLICANT: Bowman, Cindy G.
APPLICANT: Dawson, John L.
APPLICANT: Dunder, Erik M.
APPLICANT: Pace, Gary M.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6018104artis Corporation
STREET: Patent & Trademark Dept., 520 White Plains
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-9005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,595A
FILING DATE: 02-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/951,715
FILING DATE: 25-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40403
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3
TELEPHONE: (919)541-8582
TELEFAX: (919)541-8689
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3468 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..3468
OTHER INFORMATION: /product= "Full-length pure maize
OTHER INFORMATION: optimized synthetic Bt"

OTHER INFORMATION: /note="Disclosed in Figure 3 as syn1t.mze"
US-08-459-595A-2

Query Match 3.3%; Score 33.4; DB 3; Length 3468;
Best Local Similarity 45.8%; Pred. No. 3.1;
Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 27 tcaactatccaccgagacatcttccccctgctcaccagttccgcgtgtaagaagg 86
DB 2468 TCAGATCAAGACCGCCGACCGCCGCGGCGGCAAGTCTGAGAGAGA 2527
QY 87 gcccaacctcagttggtccatggttcaggtgtaatgctcctcctaagtaaccgcctccgc 146
DB 2528 AGCCCTGCTGGGCGAGCGCCCTGGCCGCGTGAAGCGCCGAGAGAGAGTGGCGGACA 2587
QY 147 ccagagagagatgagagagagatgagagagatgagagagatgagagagatgagagagatgag 206
DB 2588 AGCGGAGAGAGTGGAGTGGAG 2647
QY 207 aggcgtcagctccaccatccaccatccaccatccaccatccaccatccaccatccaccatccacc 266
DB 2648 AGCGCTGCTGGAG 2707
QY 267 acggccagcc 277
DB 2708 TCACAGCGCCG 2718

RESULT 14
US-08-459-595A-4
Sequence 4, Application US/08459595A

Patent No. 6018104

GENERAL INFORMATION:

APPLICANT: Kozielec, Michael G.

APPLICANT: Desai, Nalini M.

APPLICANT: Lewis, Kelly S.

APPLICANT: Kramer, Vance C.

APPLICANT: Warren, Gregory W.

APPLICANT: Ewola, Stephen V.

APPLICANT: Crossland, Lyle D.

APPLICANT: Wright, Martha S.

APPLICANT: Merlin, Ellis J.

APPLICANT: Launis, Karen L.

APPLICANT: Rothstein, Steven J.

APPLICANT: Bowman, Cindy G.

APPLICANT: Dunder, Erik M.

APPLICANT: Pace, Gary M.

APPLICANT: Suttle, Janet L.

TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED

TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6018104artis Corporation

STREET: Patent & Trademark Dept., 520 White Plains

STREET: Rd., POB 2005

CITY: Tarrytown

STATE: New York

COUNTRY: USA

ZIP: 10591-9005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/459,595A

FILING DATE: 02-JUN-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/951,715

FILING DATE: 25-SEP-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991

ATTORNEY/AGENT INFORMATION:

NAME: Pace, Gary M.

REGISTRATION NUMBER: 40403

REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919)541-8582

TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3468 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "Synthetic DNA"

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: misc-feature

LOCATION: 1..3468

OTHER INFORMATION: /product="Full length synthetic

OTHER INFORMATION: maize optimized"

OTHER INFORMATION: /note="Disclosed in Figure 3 as syn1t.mod. This sequenc

OTHER INFORMATION: identical to flsyn1t.fin as disclosed in Figure 1."

US-08-459-595A-4

Query Match 3.3%; Score 33.4; DB 3; Length 3468;
Best Local Similarity 45.8%; Pred. No. 3.1;
Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 27 tcaactatccaccgagacatcttccccctgctcaccagttccgcgtgtaagaagg 86
DB 2468 TCAGATCAAGACCGCCGACCGCCGCGGCGGCAAGTCTGAGAGAGA 2527
QY 87 gcccaacctcagttggtccatggttcaggtgtaatgctcctcctaagtaaccgcctccgc 146
DB 2528 AGCCCTGCTGGGCGAGCGCCCTGGCCGCGTGAAGCGCCGAGAGAGAGTGGCGGACA 2587
QY 147 ccagagagagatgagagagagatgagagagatgagagagatgagagagatgagagagatgag 206
DB 2588 AGCGGAGAGAGTGGAGTGGAG 2647
QY 207 aggcgtcagctccaccatccaccatccaccatccaccatccaccatccaccatccaccatccacc 266
DB 2648 AGCGCTGCTGGAG 2707
QY 267 acggccagcc 277
DB 2708 TCACAGCGCCG 2718

RESULT 15
US-08-459-504B-2
Sequence 2, Application US/08459504B

Patent No. 6075185

GENERAL INFORMATION:

APPLICANT: Kozielec, Michael G.

APPLICANT: Desai, Nalini M.

APPLICANT: Lewis, Kelly S.

APPLICANT: Kramer, Vance C.

APPLICANT: Warren, Gregory W.

APPLICANT: Ewola, Stephen V.

APPLICANT: Crossland, Lyle D.

APPLICANT: Wright, Martha S.

APPLICANT: Merlin, Ellis J.

APPLICANT: Launis, Karen L.

APPLICANT: Rothstein, Steven J.

APPLICANT: Bowman, Cindy G.

APPLICANT: Dunder, Erik M.

```

? APPLICANT: Pace, Gary M.
? APPLICANT: Suttle, Janet L.
? TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
? TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
? NUMBER OF SEQUENCES: 94
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: No. 6075185artis Corporation
? STREET: 3054 Cornwallis Road
? CITY: Research Triangle Park
? STATE: NC
? COUNTRY: USA
? ZIP: 27709
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/459,504B
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/459,595
? FILING DATE: 02-JUN-1995
? APPLICATION NUMBER: US 07/951,715
? FILING DATE: 25-SEP-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/772,027
? FILING DATE: 04-OCT-1991
? ATTORNEY/AGENT INFORMATION:
? NAME: Weigs, J. Timothy
? REGISTRATION NUMBER: 38,241
? REFERENCE/DOCKET NUMBER: CGC1577/CIP/DIV
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (919)541-8587
? TELEFAX: (919)541-8689
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3468 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "Synthetic DNA"
? HYPOTHETICAL: NO
? FEATURE:
? NAME/KEY: misc.feature
? LOCATION: 1..3468
? OTHER INFORMATION: /product= "Full-length pure maize
? OTHER INFORMATION: optimized synthetic Bt"
? OTHER INFORMATION: /note= "Disclosed in Figure 3 as syn1T.mze"
US-08-459-504B-2

```

```

Query Match      3.3%; Score 33.4; DB 3; Length 3468;
Best Local Similarity 45.8%; Pred. No. 3.1;
Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```

```

QY 27 tcaactcatccaccgagatcttcccccgtgtaaccagttccgtgtaagaagagg 86
DB 2468 TCAGATCAAGACCAAGAGCCAGCCCGCTGGGCAACTGAGATTCTTGAGGAGA 2527
QY 87 gcccaacctcagtggtgccatggttcaagggtgaatgcctcctcaagtaccagtcgc 146
DB 2528 AGCCCTGTGTGGCGAGGCCCTGCGCGCTGAAGCGCGCGGAGAAAGTGGCGGACA 2587
QY 147 ccaggaggagtgagcagaggatgcatattacttgcgaatcctgaggaaatcattgtg 206
DB 2588 AGCGCAGAGAGCTGAGTGGAGACCAACATCGTGTACAGAGAGCCCAAGAGAGCGCTGG 2647
QY 207 aggcgtgcagctcccaacttccagcagagcgtgcagaggtacagagagtgagtgagcag 266
DB 2648 ACCGCGTGTTCGTAACAGCCAGTACAGCCGCTGCAAGGCCGACCAACATCGCATGA 2707

```

```

QY 267 acggcccaacc 277
DB 2708 TCCAGCGCGCGCC 2718

```

```

Search completed: October 28, 2001, 21:48:02
Job time: 2650 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 19:42:27 ; Search time 1157.4 Seconds
(without alignments)
8175.485 Million cell updates/sec

Title: SEQINS_COPY_1140_2140
Perfect score: 1001
Sequence: 1 tcgcacgcacacaaatcca.....aagccatcagcgtgggagatg 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Oy	600	TGCACACCTGTCGTGGAAAGTGNCTATTCGCGGGGACACCATTGCCGCTCGAGGCTCGG	659
Oy	883	tccggatgtggaagaagatgccacccttcctgatcatgaaggcaacctggaagtgttgg	942
Db	660	TCCGGATGTGGGGAACCATGCCA-CCCTCCTGATTCATCAAGAAGC--ACCTGGAAGATGGTGACGG	716
Oy	943	aagagaagacagt-ggaaagacacacagcacaagatgccaatgcatacgctgagcg	998
Db	717	AAGGAGAAAGACTGGSGACACAAGCACCGCCGACGTTCCCAAGGCAITCAGCGGGG	773
RESULT	2		
LOCUS	AK004136		
DEFINITION	Mus musculus 18 days embryo cDNA, RIKEN full-length enriched library, clone:1110037C17, full insert sequence.		
ACCESSION	AK004136		
VERSION	AK004136.1	GI:12835201	
KEYWORDS	CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA, clone:1110037C17.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci,P. and Hayashizaki,Y		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Methods Enzymol. 303, 19-44 (1999)		
REFERENCE	2 (sites)		
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
REFERENCE	3 (sites)		
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoke,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Kluchel,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Hatada,A., Yamamoto,R., Matsunoto,H., Sekiguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.		
TITLE	RIKEN integrated sequence analysts (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
REFERENCE	4 (sites)		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and PANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
MEDLINE	5 (bases 1 to 1510)		
REFERENCE	Aadachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Futuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirokawa,T., Horii,F., Imoto,I.K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawaji,T., Kojima,Y., Kono,H., Koude,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ono,M., Okazaki,Y., Oshida,T., Owa,C., Saio,H., Saio,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)		

COMMENT	
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.	
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAAGCGCCGCACATCAGATTCTTTTCTTTTTTTTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAGACCAGAGCTCAATTAATTATTAACCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.	
FEATURES-	
source	Location/Qualifiers 1..1510 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /db_xref="MGI:106352" /db_xref="MGD:MGI:1900733" /clone="1110037C17" /cdonc_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="18 days embryo" <1..1308 /note="putative" /codon_start=1 /protein_id="BAB23185.1" /db_xref="GI:12835202" /translation="SSFSYKSEEGSTLSPVTRGECLILKYLRPRREMRQRTLLDCMD EPIAEALIEPSFOESVEERYEKKNNOENPAPEKSQPELYELTGSAIPKKTINVSST LVNLSIPKSVLLDGESTFGOLRNHGQOIDRWLCSTFAVSHLIADHDTGLNTLL QREHALASLGKFPQPLVAPQLRAWLDQYNHNCOEIIHAVSMIPAKCIQKAENVN TTERLISLTLEETDEEFOTCEPOTCLVRHCNAFGCALVHSSGMKVYSGDTPCEALYM GKDALITLHEATLEDGEDEEBAVERKTHSTSOAINVGARMMAEFTIMNHFORAKIPL ESPDENRKVGIARDHKMKVCGRDPPTVKILPKALKFAGIIEEMVERERRELRLVRA ALTIQQADSPEDEHPQOKRAHTDEPHSPQSKESVANITLGAR"
CDS	polyA_signal 1488..1493 /note="putative" polyA_site 1510 /note="putative"
BASE COUNT	373 a 412 c 414 g 311 t
ORIGIN	
Query Match Best Local Similarity 83.5%; Score 675.8; DB 192; Length 1510; Matches 781; Conservative 0; Mismatches 157; Indels 1; Gaps 1;	
OY	63 ccagttccgctgtgaagaaggagggccccaccctcaagtgtgcccacatgttcagggtgaat 122
Dd	2 CCAGCTTCTATAGTAGAAGAGAGGAAGGGTCACCCCTCACGCTGCCAAAGTTCGGGGTGAAT 61
OY	123 gccctccctaattaccagctcgctcccccaggaggaggttgycagaaagatgccattactt 182
Dd	62 GCCCTCTCAACTATACAGCTCCGCCCAAAGAAAGATGGCAAGAGGTATACCACTCGACT 121
OY	183 gcaatccctgagaattacttagttgagggctgcagatctcccaacttccagacaagcgtc 242
Dd	122 GCATTAATGTAATAATTCCTACTAGAGCGCTTTGGAGCTCCCATGTTTCCAGGAGAGTG 181
OY	243 aggaatacagaaggagtgacagagacggccccagccccagcagagaaaaagaagtaagacc 302
Dd	182 AGGAATATCGGAAGACGTCGAGAAAACCCAGCCCGCAGAGAAAGAACCAATATC 241
OY	303 cagaatacatcttcctcctggaaacagggtctgcacatcccgatgaagattcgaaatgta 362
Dd	242 CTGAATATGTTCTTCCTGGGTACGGGGGTTCTGCCATCCCAATGAATCCGAATATGTCATT 301
OY	363 ccaacctgtccaataagaccgccgacaagctctcgtcatctcgtgagatgtggtgaggacaat 422
Dd	302 CCAACTCTGTCACCTTAAGCCCTTACAAGTATAGTGGCTCTGGATTTGGAGAAAGCACTT 361

FEATURES	source
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 735)
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/ .
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov Plate: L10M841 row: j column: 23 High quality sequence stop: 732. Location/Qualifiers 1..735
RESULT	3
BE902696	735 bp mRNA EST 28-SEP-2000
LOCUS	60167733931 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959926 5',
DEFINITION	mRNA sequence.
ACCESSION	BE902696
VERSION	BE902696.1 GI:10393148
KEYWORDS	EST.
SOURCE	human.
Db	901 ACACACACACACCTCCAGGCTATTATGTGGGATG 939
Qy	963 acacacagacaaagctccaaagccatcagctggtggatg 1001
Db	841 ACCCTCCGTATACATGAAAGCCACTCTGGAGATGCTTGGAAAGAGAAAGAGAGAG 900
Qy	903 accctccgtatatacatgaaagccacctggaagaatgttttgaagaaggaacagtgaaag 962
Db	781 GTCGCTACTCTGGGGGATACCATGCGCTGTAGGCTGTGTCACAGATGGGAAAGATGCC 840
Qy	843 gtggtctatctccggggaacacatgcacctgcagagctctgtgctcgatctggggaagatgcc 902
Db	721 CTGTGATGCGGACATGCAACATATGCTTTGGCTGTGACCTGGTACATTCATCTGGCTGAAA 780
Qy	783 ctggtggtggtgacatgaaacatgctgttggctgtgcctgtggtgacacctctggttgaaa 842
Db	661 TTGGAAAGCGTATGATGAGCTTCTGTGGAAACATGTGACTTGAAGAAATTTCAACACCTGCG 720
Qy	723 gtggaagaatgtatcagttcagttcgttgcgaacaatgtgatttgaagaagttcagaacctgt 782
Db	601 CACGTCATATGATGATTCCTGCGCAATATGCTTACGAAGAAGGGGAGAGGCTCCATACACCA 660
Qy	663 cacatcaatataatctccctgcgcaaatgaccttaaggaagaaggtgtgatactccagtcctga 722
Db	541 CCTACCCAGCTGATAGGGCTGTGCTGAGCATATACAAACCACTCCACAGAGATTTCTCAC 600
Qy	603 cccaaacagctccaaagccttggtctccagatgacacaaacagcagtcgaggggtctgcac 662
Db	481 CAGAGAGAGCATATGCTTGGCATCTCTGGGGAAACCCCTTCACGCCCTTGCTGTGGTGGCT 540
Qy	543 cagagaagaagcgccttggtgaatcttttggaaagcgccttcaaccttgcattgtgttgc 602
Db	422 CTGTGTTGTGTGCCACCT-GCAGCGCAGACACACACAGGGCTTGCTGATATCTTCTG 480
Qy	483 ctgttctgttccacccttgacgacgaatataccacacacagccttgcacaagtatctgtg 542
Db	362 TTGGGCACTTTGGCGCTATTACGACAGCAAAATAGACCGAGTCTTATGACCTTCAGG 421
Qy	423 ttgggcaacttgcgcgtatataaggaagccaggtggtgacaggttcttggaaacctgtg 482

			/clone.lib="NH.MGC.21"	
			/tissue_type="Choriocarcinoma"	
			/lab_host="DH10B (phage-resistant)"	
			/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;	
			Site_2: EcoRI; cDNA made by oligo-dT priming.	
			Directionally cloned into EcoRI/XhoI sites using the	
			following 5' adaptor: GGCACGAG(G). Size-selected >500bp	
			for average insert size 1.8kb. Library constructed by	
			Ling Hong in the Laboratory of Gerald M. Rubin (University	
			of California, Berkeley) using ZAP-cDNA synthesis kit	
			(Stratagene) and Superscript II RT (Life Technologies)."	
BASE COUNT	168 a	206 c	203 g	157 t
ORIGIN				1 others
Query Match	64.4%; Score 644.6;	DB 141;	Length 735;	
Best Local Similarity	97.3%;	Pred. No. 6.3e-166;		
Matches 719;	Conservative 0;	Mismatches 14;	Indels 6;	Gaps 6;
OY	51	tcccctcgtccacagtttcgcgtgtaagaaggagccccaccctcagtgtagccattgg	110	
Db	1	TCCCCCCTCCTCACCGTTTCCGGCTGTAAAGAAGGAGGCCCCACCCTCATGTGTGCCCAATG	60	
OY	111	ttcaagggtgaatgctctctccaagtaccagcttcgfcgccaaaggaaagtgaggaaagtg	170	
Db	61	TTCAGAGTAGAATCGCTCCTCAAGTACCAAGCTCCGTCACAGAGGAAGTGCGCAGAGGATG	120	
OY	171	ccattatlaact-gcaatccbtggaatatcatagtttgaggcgctcagctcccaattc	229	
Db	121	CCATTATTACTTNGCAATTCCTGAGGAATTCATAGTTGAGGCCCTCAGCTTCCCACATTC	180	
OY	230	cagcagagcgtgcaagagatacagaagagatgctgcagacaagcgccaccgcccaagagaaa	289	
Db	181	CAGCAGAGCGTSCAGAGATACAGAGAGAGTGGCGCAGAGCGGCCACCCACACAAGAAA	240	
OY	290	agaagtcagtaacccagaatatcatcttctcttggaaagaaggtctgcataccgatgaagatt	349	
Db	241	AGAACTCATGATCCCAAGAAATCATCTTCCCTTGAAACAGAGGTCTGCATCCCGATGAAGATT	300	
OY	350	cgaaatgtcagtgccacactgttaacataagccccgaacaagttctctctaattgacctgt	409	
Db	301	CGAAATGTCTAGTGCACACTTGTCAACATTAAGCCCCGACAGCTCTCTGCTACTGGACTGT	360	
OY	410	gctgagggcacatttgygcagctgtgcgtcatltaacggagacaggtgtagaacaggtctcgt	469	
Db	361	GGTAGGGGACATTTGGGAGGCTGTGCCGTATPACGGAGACACAGTGGACAGGGGTCTTG	420	
OY	470	ggacccctgctcgtcgtgttttgttcccaactgtgcagcgagataccacaagggtctgac	529	
Db	421	GGCACCTCGGCGCTGCTGT-TTGTGTCCCACT-GCACGCAGATTCACACACGGGCTTGCC	478	
OY	530	aagtatcttgcgtgagagagaagacgcgctltgcatctlttggaaagacgcgtccaccctt	589	
Db	479	AAGTATCTTGTGTCAGAGAGGAACGGCCTTGGCATC-TTGGAAAAAGCGCTTCACCCCTT	537	
OY	590	gctggtgtgttgcggcccaaccacagctcaaaagcctgagctccagcagtaaccaaacagtgcoa	649	
Db	538	GCTGTGTGTGGCCCCAACACACACTCAAAGCGCTGGCTCCAGCAGTACCAACAACATGTGCCA	597	
OY	650	ggaagtlcctgcaaccaatcagtaagattcctgcgcaaatgtccttcaggaagggtctgagat	709	
Db	598	GGAGGTCTCTGCACCAACATCAGTATATTCTTG-CAAATGCCCTTTCAGGAAGGGCGTGAGAT	656	
OY	710	ctccagctctgtgaugggaaagatltagttagtgcgtgtgttggaacaatgtattt-ggaag	768	
Db	657	CTCCAGCTCTCTGAGAGGAAGATTGATCATGTCGCTGTTCGCAACATGATGATTTGGGAAG	716	
OY	769	aattcacagacctgtcgtgt	787	
Db	717	AGTTTCAGAGACTGTCTGTT	735	

BF525432	LOCUS	BF525432	947 bp	mRNA	EST	11-DEC-2000
DEFINITION		602069517/1	NCI_CGAP_Brn64	Homo sapiens	CDNA clone	IMAGE:4212365
ACCESSION		BF525432				
VERSION		BF525432.1	GI:11612793			
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS		1 (bases 1 to 947)				
TITLE		NIH-MGC http://mgc.nci.nih.gov/.				
JOURNAL		National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT		Unpublished (1999)				
		Contact: Robert Strausberg, Ph.D.				
		Email: cgapbs-remail.nih.gov				
		Tissue Procurement: David N. Louis, M.D.				
		CDNA Library Preparation: Life Technologies, Inc.				
		CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
		DNA Sequencing by: Incyte Genomics, Inc.				
		Clone distribution: MGC clone distribution information can be				
		found through the I.M.A.G.E. Consortium/LLNL at:				
		http://image.llnl.gov				
		Plate: L1AM9783 row: a column: 06				
		High quality sequence stop: 655.				
FEATURES		Location/Qualifiers				
source		1..947				
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		/clone="IMAGE:4212365"				
		/clone_1lb="NCI_CGAP_Brn64"				
		/tissue_type="gliblastoma with EGFR amplification"				
		/lab_host="DH10B (T1 phage-resistant)"				
		/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1; Notif: Site 2: SalI; Cloned unidirectionally; Primer: Oligo dT. Average insert size 1.57 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."				
BASE COUNT		224 a 257 c 289 g 177 t				
ORIGIN						
Query Match		63.6%; Score 636.8; DB 150; Length 947;				
Best Local Similarity		96.5%; Pred. No. 9.4e-164;				
Matches 737; Conservative		0; Mismatches 17; Indels 10; Gaps 8;				
QY	241	gcagagagtcacgaaggaatgctgcagagagcgccaccccccagagagagaaagaatcaca	300			
Db	1	GCAGAGAGTACAGAGAGAGTGCCTCCAGAGAGCGCCACGCCAGCCACGAGAGAAAGAGTCAAGTA	60			
QY	301	cccagaatcatctctctctgtgacacagggctctgcacatcccgatgaagatcgaatgtcag	360			
Db	61	CCCAAGAAATTCATCTTCCTTGGAACAGGGGTCCATCCCGATGAAGATTGCAAAATGTGAC	120			
QY	361	tggcacacttgtcaacataaagccccgaacagctctctgtactgtgactgtgtgtgagggc	420			
Db	121	TGCCACACTTGTCAACAATAAGCCCGAACAAGCTCTGTCTACTGAGACTGTGTGAGAGGCGAC	180			
QY	421	attgggagcgtgtgctgcatattacggagacaggttgacaaggtgtcctgtggagacctgac	480			
Db	181	ATTGGGAGAGCTGTGCTCCTCATTTACGGAGACCAAGGTGACAGAGGTTCTTGAGACCTTGAC	240			
QY	481	tgtctgtgttgttgcacaccttgacacagatcatcacacaacagagcttgcacaatcatcttc	540			
Db	241	TGCTGTG-TTGTGTCCACACT-GCACGCGAGATCACACAGGGCTTGGCAATATCTTGC	298			
QY	541	tgcagagagaaacgcgcttgcacatcttctgggaagacgcgcttcaaccttgcgtgtgtgtg	600			
Db	299	TGCAGAGAGAAACGCGCTTGGGATCTTTGGGAAACCGCTTCAACCTTGTGCTGTGTGTG	358			
QY	601	cccccaaccagctcaaaagcctgtgtcttcagagcaggtaccacaacacaggtccagaggtctcgc	660			
Db	359	CCCCCAACCGAGTCAAAAGCTGTGGCTCCAGACATATCACCAACAGGAGGAGGTTCTTGC	418			

OY	661	aacataatgaatgatattcttcgcgcaaatggccttcaaggaaagggcttgagatctccagtcctcg	720
Dd	419	ACCACATCAGTATATATTCTTCGCCAAATGCGCTTCAGGAAGGGCGCTGAGATTTCCAGTCTCG	478
OY	721	cagtggaaaagaattgatcatcagttcgtctgttcggacaatctgtgatttggaaagaagttccaacct	780
Dd	479	CAGTGAAAGAATGATCATGAGTTCGCGTGTTCGGAACAATGTGA-TTGGAAAGATTTCAAGACCT	537
OY	781	gtctggtgcccactgtgcaagaatgctgttggctgtgtgcctcgtgtgacacctcgtgttga	840
Dd	538	GCTGTGTCGGCACCTGCAGGATCG- TTGGCTGTGGCT-GTGCACACCTCTGGCTGGTA	595
OY	841	aagtgttatctccggggagaccatgccccgcgcagagctgtgtgccgga---tggggaaa	897
Dd	556	AAGTGTCTATTCCGGGGACACCATGCCCTTCGAGCGCTGTGTTCCGACTTGGGGAAAG	655
OY	898	atgccaccctctgatatacatgaagccacccttgaagaatgttgttgaagaagaaagcagtag	957
Dd	656	ATGCCAGCGCTCTGATCATGATGAACCCACCTCGGAGATGG-TTGGAAAGAGGAAGCAGTGG	714
OY	958	aaaaagacacagcacaaagctgccaaaccatcagctggctggagatg	1001
Dd	715	AAAAGCCACGACGACAA-GTCCCAGCATCAAGCGGGGGGATG	757
RESULT	5		
LOCUS	BE795820	761 bp	mRNA
DEFINITION	60159085F01 NIH_MGC_7	Homo sapiens cdna clone IMAGE:3945085	5', mRNA sequence.
ACCESSION	BE795820		
VERSION	BE795820.1	GI:10217018	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabs-remail.nih.gov		
	Tissue Procurement: DCTD/DRP		
	CDNA Library Preparation: Ling Hong/Rubin Laboratory		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov plate: LICMB02 row: P column: 14		
	High quality sequence stop: 761.		
FEATURES	Location/Qualifiers		
SOURCE	1..761		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3945085"		
	/clone_id="NIH_MGC_7"		
	/tissue_type="small cell carcinoma"		
	/cell_line="MGC3"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: lung; Vector: pORF7; Site_1: XhoI; site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Size-selected >50bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."		
BASE COUNT	172 a	206 c	219 g
ORIGIN			164 t

Query Match 63.6%; Score 636.4; DB 140; Length 761;
 Best Local Similarity 99.5%; Pred. No. 1.1e-163;
 Matches 659; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 340 gatgaagatcgaanaatgctcagctgacacactgttcacaaataagcccgagacagctctgct 399
 Db 1 GATGAAGATCGAATGACAGTGCACACTGTGTCAATAGCCCCGACAGCTCTGCT 60

QY 400 actgagctgtgtgagggacacatttggcagcgtgtgcgcgtcaatacggagacaggtgga 459
 Db 61 ACTGACAGTGTGTGAGGGACAGTGTGGCAGCTGTGCGTCAATACGAGACAGAGTGA 120

QY 460 cagggtctctggcaaccctggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 519
 Db 121 CAGGCTCTTGCGACACCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 179

QY 520 cggagcttgcgaagatctctgtcgcagagagagacgacgcttggcatcttgggaaagccgc 579
 Db 180 CGGCTTGCAGATATCTGTGCTGAGAGAGAACGGCCTTGGCATCTTTGGGAAAGCCGC 239

QY 580 ttcaacctgt 639
 Db 240 TTCACCCCTTGT 299

QY 640 accagttgcagagagctctgtcacacacatgatacttctgtgcgaatgcttcaggag 699
 Db 300 ACCAGTGCAGAGAGCTCTGTGACACCATGATGATGATGATGATGATGATGATGATGATG 359

QY 700 gggctgagatctccagctctgcagtgagagagatgatcagctgtgtgtgtgtgtgtgtgt 759
 Db 360 GGGCTGAGATCTCCAGCTCTGAGTGGAAGATGATGATGATGATGATGATGATGATGATG 419

QY 760 attggaagagttcagaacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 819
 Db 420 A-TTGGAGAGTTTCAGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 478

QY 820 tgggtcacaccttgcgt 879
 Db 479 TGGTGCACACCTGT 538

QY 880 tggctcgaatggggaagaatgccaccctcctgatatcagaaagccaccctgtgagaatgtgt 939
 Db 539 TGGTCCGATGGGGAAGATGACACCTCTGTATGATGAAACCAACCTCGGAAGATGCT 598

QY 940 tgggaagagagagcagtggaagagacacacagcacaagtcaccaagccatcagcgtgtgt 999
 Db 599 TGGAGAGAGAGCACTGGAAGAAAGACACAGCACAACGTCACCAATCAGCGTGGGA 658

QY 1000 tg 1001
 Db 659 TG 660

RESULT 6
 LOCUS BE260495 676 bp mRNA EST 26-OCT-2000
 DEFINITION 601150702P1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
 mRNA sequence.
 ACCESSION BE260495
 VERSION BE260495.1 GI:9131807
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 676)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs.femail.nih.gov
 Tissue Procurement: ATCC
 COMMENT cDNA Library Preparation: Ling Hong/Rudin Laboratory

cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: L10M176 row: d column: 01
 High quality sequence stop: 672.
 Location/Qualifiers
 source
 1. 676
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3503184"
 /clone_lib="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

BASE COUNT 154 a 207 c 176 g 139 t
 ORIGIN

Query Match 61.7%; Score 618; DB 165; Length 676;
 Best Local Similarity 99.8%; Pred. No. 1.2e-158;
 Matches 629; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 tggcgcacaaagatccaaccagctcaactatccaccggagatctccctgt 60
 Db 48 TGGCAGCCACCAAGATTCACACCGCTCAACTCATCCACCGGAGATCTCCCTGCT 107

QY 61 caaccagttccgctgtaagaagagggcccaaccctcagtggtccatggtcaaggtga 120
 Db 108 CACCAATTCCTGTAAGAAAGAGGGCCCCACCCCTCACTGTGCCATGTGTGAGGCTGA 167

QY 121 atgctcctcaagtaaccagctccgtccacagagagaggtgcagagagatgcatattac 180
 Db 168 ATGCCCTCCCAAGTACACACCTCCGTCCAGAGAGAGTGGCAGAGGATGCCATTATTC 227

QY 181 ttgcaatccttgaggatattatggtgagcgctgcagcttcccaattccagagagct 240
 Db 228 TTGCAATCCTTGAGGAATTAATGATGAGCGCTGCGACTTCCCACTTCACGACAGCG 287

QY 241 ggaagagtaagaagagagtgagcagagagccagccagcagagagaaagaagtca 300
 Db 288 GCAGAGATACAGAGAGAGTGGCCAGAGACGGCCAGCCCAAGCAGAGAAAGATCAGTA 347

QY 301 ccagagaatcatcttcccttggaaacagggctcgcacatcccgatgaagatcgaaatgtcag 360
 Db 348 CCAGAAATCATCTCTTGGAAACAGGAGGTGTCATCCGATCCGATGAAGATTCGAAATGTCA 407

QY 361 tggcaactgttcaacataagcccgacacgctctcgtctactcagatgtgtgtgtgtgtgtgt 420
 Db 408 TGCACACTTGTCAACATTAAGCCCGACACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 467

QY 421 atttggcagcgtgtgcgcgtcatcagagagacaggtgtgacagaggtcctgtggacccctgtgc 480
 Db 468 ATTTGGGCGAGCTGTGCGCTCATTTACGGAACCAAGGTGGACAGGGTCTTGCGACCTGGC 527

QY 481 tgcgt 540
 Db 528 TCGT 586

QY 541 tgcagagaaagcgcgt 600
 Db 587 TGCAGAGAGAACCGCGCTTGGCATCTTTGGGAAAGCGCGCTTCAACCTTGTGTGTGTGTG 646

QY 601 cccccaacagctcaagcgt 630
 Db 647 CCCCCACACGCTCAAAAGCTGTGCTCAGC 676

```

RESULT 7
LOCUS BE382353 692 bp mRNA EST 21-JUL-2000
DEFINITION 601298656F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3629028 5',
            mRNA sequence.
ACCESSION BE382353
VERSION BE382353.1 GI:9327718
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 692)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LNC314 row: o column: 13
High quality sequence stop: 600.
FEATURES
    source
        1..692
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3629028"
            /clone_lib="NIH_MGC_19"
            /tissue_type="neuroblastoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: brain; Vector: pORF7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCACGAG(G). Library constructed by Ling Hong
            in the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-CDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH-MGC Library."
BASE COUNT 152 a 193 c 197 g 150 t
ORIGIN
Query Match 59.9%; Score 599.8; DB 166; Length 692;
Best Local Similarity 98.5%; Pred. No. 1.2e-153;
Matches 637; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY 336 tcccgatgaagatcgaatgctcagtcagtcacattgcaacaagcccgacagctc 395
    |||||||
DB 1 TCCCGATGAAGATTGCAAAATGTCAGTCCACACTGTCACATTAAGCCCGACACGTCTC 60
    |||||||

QY 396 tctactggaacttggtgaagggacacatttggaagctgtgcgtcattacgaagacagg 455
    |||||||
DB 61 TGTACTGGAAGTGTGTGAGAGGACCAATTTGGGACGCTGTCCGTCATTACGAGACACAG 120
    |||||||

QY 456 tggacaaggctcgtggaacccctgctgtgtgtttgttccacactggaagcagatcac 515
    |||||||
DB 121 TGGACAGAGGCTGTGGGACACCTGCTGTGTGTGTGTCACACT-GCACGCGATCAC 179
    |||||||

QY 516 caaagggtctgcaagatattctgtgcaagaagaagcgacctgtgcatcttgggaag 575
    |||||||
DB 180 CACACGGGCTTGGCAATATCTTGTCAGAGAGAACGCCCTTGGCATCTTGGGAAG 239
    |||||||

QY 576 ccgctcaccttgcgtggtgttgccccaacacagctcaagcctgtccacagctac 635
    |||||||
DB 240 CCGCTTACCCCTTTGCTGGTGTTGGTGGCCCAACCAAGCTCAAGCTGCTCCAGCAATAC 299
    |||||||

QY 636 cacaaccagtcagagagctctgacacacatcagatgatctctgccaatgcttcag 695

```

```

|||||
DB 300 CACACACAGTGCACAGAGGATCTGCACACATCAGTATGATTCCTGCCAATGCTTCAG 359
    |||||||

QY 696 gaagggtcagaatctcccaatctcctgcagttgaaagattgactgctgctggcagaca 755
    |||||||
DB 360 GAAGGGGCTGAGATCTCCATCTCGAGTGAAGATTGATGATTCGCTGTGGCAACA 419
    |||||||

QY 756 tctgattgaaagattcagacctgtctgtgcgcagctcgaagcagatgctgtgctgt 815
    |||||||
DB 420 TGTGATTGGAAGAGTTTCAGACCTGTCTGTGTGGGACACGCAAGCATGGCTTGGCTGT 479
    |||||||

QY 816 ggcgtgtgcacacctctgctggaagtgctctatctcggggagaccatgacctgcag 875
    |||||||
DB 480 GCGCTGTGCACACCTCTGCGTGAAGTGTCTATTCGGGGACACCATGCTCGCAG 539
    |||||||

QY 876 gctctgctcggatgggaaagatgcccacctctgataatgaagccacctgggaat 935
    |||||||
DB 540 GCTCTGCTCGGATGGGAAAGATGCCACCTCTGTATGATGAAGCCA-CTGGGAAGAT 598
    |||||||

QY 936 gatttgaagaggaagcagtcgtgaaagacacacagcacaacgctccca 982
    |||||||
DB 599 GCTTTGGCAGAGGAGCGCTGG-AAAGACACACAGCAGCTGCCAA 644
    |||||||

RESULT 8
LOCUS BG519751 884 bp mRNA EST 02-APR-2001
DEFINITION 602578902F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
            mRNA sequence.
ACCESSION BG519751
VERSION BG519751.1 GI:13515513
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 884)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LNCM76 row: d column: 01
High quality sequence stop: 859.
FEATURES
    source
        1..884
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3503184"
            /clone_lib="NIH_MGC_19"
            /tissue_type="neuroblastoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: brain; Vector: pORF7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCACGAG(G). Library constructed by Ling Hong
            in the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-CDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH-MGC Library."
BASE COUNT 199 a 251 c 249 g 185 t
ORIGIN
Query Match 57.6%; Score 576.4; DB 154; Length 884;
Best Local Similarity 96.3%; Pred. No. 3.4e-147;
Matches 622; Conservative 0; Mismatches 21; Indels 3; Gaps 3;

```



```

ACCESSION   BG335963
VERSION     BG335963.1
KEYWORDS    GI:13142401
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 614)
  NIH-MGC http://mgs.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapds-r@mail.nih.gov
  Tissue Procurement: ATCC
  CDNA Library Preparation: Ling Hong/Rubin Laboratory
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LNL at:
  http://image.lnl.gov
  Plate: LCM1222 row: 0 column: 02
  High quality sequence stop: 614.
  Location/Qualifiers
    source
      1..614
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4542553"
        /clone_1ib="NIH_MGC.21"
        /tissue_type="choriocarcinoma"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
        Site_2: EcoRI; CDNA made by oligo-dT priming.
        Directionally cloned into EcoRI/XhoI sites using the
        following 5' adaptor: GGCAAGAG(G). Size-selected >500bp
        for average insert size 1.8kb. Library constructed by
        Ling Hong in the laboratory of Gerald M. Rubin (University
        of California, Berkeley) using ZAP-CDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      135 a      169 c      178 g      132 t
Query Match      56.7%; Score 567.6; DB 152; Length 614;
Best Local Similarity 97.4%; Pred. No. 7.8e-145;
Matches 598; Conservative 0; Mismatches 14; Indels 2; Gaps 2;
```

```

QY 366 cactgtgcaacaataagcccgacacgctctgtctactgactggtggtgagggacatttg 425
    |||||||
DB 2 CACTGTGCAACATAAGCCCGACACGTCCTGTCTACTGTGAGCTGTGAGGGCACATTTG 61
QY 426 ggcagctgtgcgtcattacaggaacacaggtgagcaggtctctgtggcacccctgtgctg 485
    |||||||
DB 62 GGCAGCTGTGCGTCATTACGAGACACAGGTGAGGAGGTCCTGTGGCACCCCTGTGCTG 121
QY 486 tttgtgtccacactgagcagcagatcacacacacgggcttgcgaatlatctgtcgaag 545
    |||||||
DB 122 TGTGTTGTGTCACACT-GCACGAGATCACACACAGGGGCTGCAAGATATCTCTGCGAG 180
QY 546 aaggaagcgcttggatatttgggaaagccgcttcaaccttgcgtggtgtgccc 605
    |||||||
DB 181 AGAGAAAGCGCCCTGGCATCTTTGGGAAGCCGCTTCACCTTTGTGTGTTGCCCC 240
QY 606 aaccagctcaaaagctgtgctcacaagatcacacaacacagtgccaggaaggtctctgacac 665
    |||||||
DB 241 AACCAAGCTCAAAAGCTGTGCTCCAGCATACCAACAGTGTCCAGAGGTCTCTGACACAC 300
QY 666 atcagatgatctccgtcnaatgcttcaggaagggctgagatctccagtcctgcagtg 725
    |||||||
DB 301 ATCAGATGATTCCTCGCAATGCTCTCAGGAAGGGCTGAGATCTCCAGTCTCGACAGTG 360
QY 726 gaagaatgatcagtgctgctgttggaacatgtgatttggagaaggttccagactgtctg 785
    |||||||
DB 361 GAAAGATGTGATCAAGTGTGCTTGCGAAGCATGTGATTTGGAAAGATTTCACACCTGTCTG 420
    ..
```

```

QY 786 gtgcgcactgcaagcattgcgttgcgtgtgcgtgtgcacacactgtgctgaaagt 845
    |||||||
DB 421 GTGCGGCACTGCACAGCATGCTTTGGCTGTGCTGTGCTGTCACACTCTGCTGGAAGTG 480
QY 846 gtctattccgggaacaccatgccttcgagcgtctgtgtccgattggg-aaagatgccac 904
    |||||||
DB 481 GTCTATTCCGGGACACCATCCCTGCGAGGCTGTGCTCGGATGCGGAAAGATGCCAC 540
QY 905 cctctatcatataagccaccctgggaagatggttggagaaggaagcagtggaagac 964
    |||||||
DB 541 CCTCTATATGATGAAGCCACCTGTGAAGATGTTGGAAGAGAGAGTGTGACAAAGAC 600
QY 965 aacagacacaacgt 978
    |||||||
DB 601 ACACAGCAACAACGT 614
RESULT 11
BE383336      664 bp      mRNA      EST      21-JUL-2000
LOCUS        601298249F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5',
DEFINITION   mRNA sequence.
ACCESSION   BE383336
VERSION     BE383336.1
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 664)
  NIH-MGC http://mgs.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapds-r@mail.nih.gov
  Tissue Procurement: ATCC
  CDNA Library Preparation: Ling Hong/Rubin Laboratory
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
  Plate: LCM313 row: a column: 13
  High quality sequence stop: 662.
  Location/Qualifiers
    source
      1..664
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:3628308"
        /clone_1ib="NIH_MGC_19"
        /tissue_type="neuroblastoma"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
        EcoRI; CDNA made by oligo-dT priming. Directionally
        cloned into EcoRI/XhoI sites using the following 5'
        adaptor: GGCAAGAG(G). Library constructed by Ling Hong
        in the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using ZAP-CDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies).
        Note: this is a NIH_MGC Library."
```

```

FEATURES
  source
```

```

BASE COUNT      154 a      199 c      174 g      137 t
ORIGIN
```

```

Query Match      56.4%; Score 564.8; DB 166; Length 664;
Best Local Similarity 99.2%; Pred. No. 4.7e-144;
Matches 599; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
```

```

QY 1 tcgaagcacaagaattcaaacacagctcaacctatcacaccggacatcttccctgct 60
    |||||||
DB 48 TCGAGCCACAGAGTTTCAAAACCCAGCTCAACCTATATCCACCGGACATCTTCCCTGCT 107
QY 61 caccagttccgtgttaagaagagggcccaacctcagtggtgccatggttcagggtga 120
```

Db	108	CACCAGTTTCGGCTTAAAGAAGGAGGCCCCACCCCTCAgTGtGCCATGTGCACAGGTGA	167
Oy	121	atgtcctccaagtaaccagcttcgcgtccagaaggagtggcacaggaatgcattataac	180
Db	168	ATGCTCTCTCAAGTACACAGCTCGTCGCCAGAGGGAGTGGCAGAGGAGTGCATTATTAC	227
Oy	181	ttagcaatccttgaggaattcatatagtttgaggcgctgcaagcttccccaaccttcagcagaagct	240
Db	228	TTCGCAATCCCTGAGGAATTCATATAGTTGAGGCGCTGGAGCTTCCCAACTTCACACAGAGCGT	287
Oy	241	gcagaagtaaccaagaagatgacgcagagaagcccagccccagacagaaagaagtcacta	300
Db	288	GCAGAGATCACAGAGAGAGTGGCCACGACGCCACCCACGACGAAAHAAGAATCAGTA	347
Oy	301	cccagaaatcalctctccttgagacaagggltctgcatacccgatgaagatlcgaatgacag	360
Db	348	CCCAGAAATCATCTTCCTTGGAACAAGGGTCTGCCATCCCGAATGAAGATTGSAATGTGAC	407
Oy	361	tgccacctgttcaacaataagccccgacaacgctctgtctaactgtagctgtgtgaggcac	420
Db	408	TGCCACACTTGTCAACATAAGCCCCCGACACGCTCTGTACTGGAAGTGTGTGAGAGGCAC	467
Oy	421	atttgagagcgtgbcggtcatttaaggagaccaggtltagaacaggygtccctgggaacctggc	480
Db	468	ATTTGGCAGCGCTGTCCGTCATTTACGGAGACAGAGGTGACAGGGTCTGGGACCCCTGSC	527
Oy	481	tgcctgttgttctgtcccacacttgcacagcagacatcacacacagagcttgcacaatatctgc	540
Db	528	TGCTGTGTTGTGTGCCACCT- GCACGCGAGATCACACAGGGCTTGGCAATATCTTGC	586
Oy	541	tgc-agagaagaacgcgccttgacatcctttgggaaaagccgcttacaccttgcctgtgtgt	599
Db	587	TGCAGAGAGAACCGCGCTTGCACTTTGGGAAAG-GCCTTCACCCCTTGTCTGTGTGTG	645
Oy	600	gccc g03	
Db	646	CCCC C49	
RESULT	12		
LOCUS	BE537086		
DEFINITION	BE537086 713 bp mRNA	EST	09-AUG-2000
ACCESSION	601063325F1 NIH_MGC_10 Homo sapiens CDNA clone IMAGE:3449742 5'		
VERSION	BE537086	mRNA sequence.	
KEYWORDS	BE537086.1 GI:9765731		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 713)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Incyte Genomics, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LAM8427 row: a column: 07 High quality sequence stop: 710. Location/Qualifiers		
FEATURES	source		
	1..713		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3449742"		
	/clone_id="NIH_MGC_10"		

			/cell_line="MGC36"			
			/lab_host="DHIOB"			
			/note="Organ: Cervix; Vector: PCMV-SPOrt6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5 kb. Library prepared by Life Technologies."			
BASE COUNT	166 a	187 c	204 g	155 t 1 others		
ORIGIN						
Query Match	56.2%	Score 562.6:	DB 136:	Length 713;		
Best Local Similarity	92.9%	Pred. No. 1.9e-143:				
Matches	659: Conservative	0: Mismatches	34: Indels	16: Gaps 6		
Oy	116	ggatgaatgcctcctctaagtaaccagtccgtcccagaaggagttgacagagtgatgccatt	175			
Dd	8	GGTGATAGCTCCTCCCAAGTACCAGGTCCGTCACAGSAGGAATTGGCAGATGGCATG	67			
Oy	176	attaacttgcatactctgaaagaaatlcataglttagagcgctgcaagcttccccaacgtcagag	235			
Dd	68	ATTACTTTGCAATCCCGAGGAATTCATTAGTTGAGGCGCTCACCTGCCAACTCTCCAGCAG	127			
Oy	236	accctgcagaggaataagagagaggtgcagagagcgcccgccccagcaagaagaanaat	295			
Dd	128	AcCGTGCAGAGATACAGAGGAGTGGCGCAGACGGCC--AGCCACCAAGAAGAAAAGAT	185			
Oy	236	cagtaaccagaatatcatcttcctcttgaaacaaggtctctccatccagaatgaattlccaat	355			
Dd	186	CAGTACCCAGAAATCATCTTCTTGSAACAGSGGTCTGCCATCCCATGAAATTCGAAT	245			
Oy	336	gtcagltgcacaactgtgcaacaataagccccagacaogltctgctactagtaactgtgtgag	415			
Dd	246	GTCAGTGGCACACTTGTCMAACAATAAGCCCCGACAGCTCTCTACTAGTAAGTGGTAG	305			
Oy	416	ggcacatttgggcagctgtgcccgtattacagagaccaggtgaca--gggtctcgggca	473			
Dd	306	GGCAGGTTTTGGGCGACCTGTGGCGCTATTACGAGAACAGGTGACAAAGTGTCTTGCGCA	365			
Oy	474	ccctg-gctgcctgttttgtgtccacctgycagcagacagataaccaagggcttggcaag	532			
Dd	366	CCCCTGTGCTGCTGTGTATGTGTCCCAACATGCAACGACAGATCACACACGGCTTGCCAAG	425			
Oy	533	tatctgtctgcagagagaagcgccttggc-atccttggaaaaagccg-----ctlc	582			
Dd	426	TATCTTGGTGCAGAGAGAAGACGGCCCTTGCCATCTTTGGGAAAGCCGNGNTTCAACCTTT	485			
Oy	583	acccttgcctgtgtgtgtgctgccccaacacagctcaaaagcctgtctcaagcagtaaccacaacc	642			
Dd	486	TGCTTTGGTGGTTGGCGCTTGCCCAAACAGGCTTCAAAAGCCCTGCTCCAGCATGCCAACMC	545			
Oy	643	agtgcagaagaagtcctgtgacccacatcagatgatctcttcgccaatagtccttgaagagg	702			
Dd	546	AAGTGCACAGGAAGTCTGTGACACATCATGATATATTCCTCCCAAATGCCCTTCAGGAAGGGG	605			
Oy	703	ctgagatctccagtcctctcagtgagaaaagattatcaatlgcgtctgttcgaaacatgtgat	762			
Dd	606	CTGAGATCTCCAGCTGCTGTGAGAGTGAAGATGAAATCAGTTGCGTGTGTCGAACATGTGATT	665			
Oy	763	tggagaagattcagaactgtctgtgtgtgtgagcaactcagaagcatgtcggtttg	811			
Dd	666	TGGAGAGATTTCAGACCTGTCTGTGGCGG-ACTGCAAGCATCGCTTGGG	713			
RESULT	13					
BEB19259	BEB19259	823 bp	mRNA	EST		
LOCUS	60147130P1 NIH_MGC_68	Homo sapiens	cdna clone IMAGE:3876223	5',		
DEFINITION	mRNA sequence.					
ACCESSION	BEB19259	GI:9890197				
VERSION	BEB19259.1					
KEYWORDS	EST.					
SOURCE	human					
ORGANISM	Homo sapiens					

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 823)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: DCTD/DRP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LHAM936 row: k column: 08
High quality sequence stop: 695.

FEATURES
source
1. 823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3876223"
/clone_lib="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies." Location/Qualifiers

BASE COUNT 185 a 219 c 259 g 160 t
ORIGIN

Query Match 54.1%; Score 541.4; DB 138; Length 823;
Best Local Similarity 99.3%; Pred. No. 1.3e-137;
Matches 575; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

423 ttggcagcgtgtccgtcattacgagaccaggttgacaggtctctggcacccctgctg 482
10 ttggcgacgtgtgcccgtcattacgagaccaggttgacaggtctctggcacccctg 69
483 ctgtgttctgtccacgtgacgacgacgacgacgacgacgacgacgacgacgacg 542
70 ctgtgttctgtccacgtgacgacgacgacgacgacgacgacgacgacgacgacg 127
543 cagaagagaagcgtcttgatcttgggaaagccgttacccttgcgtggtgtgccc 602
128 CAGAGAGAACGCCCTTGGCATCTTTGGGAAGCCGCTTACCCCTTGGTGTGGTGGC 187
603 cccaaccagctcaaaagccttgctcagcaataccacaaccagtgccagaggtctctg 662
188 CCCAACACAGCTCAAAAGCTGTGCTCCAGCACTACACACACAGTCCAGAGGCTCTG 247
663 cacatcagatgatctcctgcgcaaatgccttcaggaagggctgagatcctcagtc 722
248 CACATCATGATGATTCCTGCTGCAAAATGCTTACGAAAGGGGCTGAGATCTCAGT 307
723 gtagaaagattgctcgtctgctgtgctgacacatgtgatttgaaagaagttcaga 782
308 GTGGAAGAGATTGATCAGTTGCTGCTGCGAATGTCATTTGGAAGGTTTCAGACTGT 367
783 ctgtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 842
368 CTGTGTGCGGACACGACGATGCGCTTGGCTGTGCGCTGCTGCTGCTGCTGCTG 427
843 gtagctatctcggggagacacatgcccctcgagagctctggtccgagtgaggaaag 902
428 GTGTGTATATCCGGGAGACACATGCGCTGCGAGGCTCTGTCGGGATGGGGAAGA 467
903 accctctgataatgaagcaccctcgaaagatggtttggaagaaggaagcagtgaaag 962
488 ACCCTCTCTATACATGAGGACACCTCGGAAGATGTTGGAGAGAGAACAG-GGAAAG 546

Qy 963 aaacacagcacaacgtcccaagccatcagcgtgaggatg 1001
Db 547 ACACACAGCACAACGCTCCCAAGCATCAGCGTGGGATG 585

RESULT 14
LOCUS BF311926
DEFINITION 601897767F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4126864 5',
mRNA sequence.
ACCESSION BF311926
VERSION BF311926.1 GI:11259697
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE 1 (bases 1 to 915)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LHAM017 row: f column: 17
High quality sequence stop: 672.

FEATURES
source
1. 915
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4126864"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC Library."

BASE COUNT 206 a 262 c 288 g 159 t
ORIGIN

Query Match 53.5%; Score 535.4; DB 147; Length 915;
Best Local Similarity 98.0%; Pred. No. 5.9e-136;
Matches 595; Conservative 0; Mismatches 6; Indels 6; Gaps 5;

Qy 1 tcgcagcacaagaattcaaacacagctcaactcatccaccggacatcttccctgct 60
Db 68 TCGAGGCACAAAGATTCAAAACCAAGCTCAACCTCATCCCGACATCTTCCCTGCT 127
Qy 61 caccagttctcgtctgaagaagagggccaccctcagtgctgctcaggttcaaggatga 120
Db 128 CACACAGTTTCGGCTGTAGAAAGAGGGCCCAACCTCATAGTGTGCCATGTTCAAGG 187
Qy 121 atgctcctcctaagatcaccagctccgtccaggaagagtgagcagaaggaatgca 180
Db 188 ATCCCTCTCTCAATACAGCTCCGCTCCAGAGGAGGTGGCAGAGGAGATGCCATT 247
Qy 181 ttcaatcccgagaattatagttgagcgctgagcgcttcccaacttccagaagagct 240
Db 248 TTCCAAATCTCGAAGGAATCTAAGTTGAGGCGTGCAGCTTCCCAACTTCCAGAGA 307
Qy 241 gcaaggatcagaagagatcgagagagcggccagcccccagagagaagaagtcagta 300

Db 308 GCAGAGTACAGAGAGTGCAGACGG-CCAGCCCGCAGAGAGAAAGATCAGTA 366
Qy 301 cccagaatcatcttccttggagacaggtctgcacatcccgatgaagattcgaatgtcag 360
Db 367 CCCAATAATCATCTTCTTGGAAAGGGTCTCCATCCCGAATGAAGATTGAAATGTCTAG 426
Qy 361 tggcaacttgcacataaagcccgacacgctctgtctactggaactgtgtggaagggcac 420
Db 427 TGCCACACTGTGCACATATAGCCCGACACGTCTCTGCTACTGAGACTGTGGTGAAGGAC 486
Qy 421 atttggagacgtgtgcgcgtacattacgagagaccaggtgagacaggtcttggacacctgac 480
Db 487 ATTTGGGAGAGTGTGCCGTCATTACGAGAGACCAGGTGAGAGGTCCTGGGCAACCTTGCC 546
Qy 481 tgcgtgtgtgtgtgtccacacttgcacgacagatcaccaacaggtgtgc-aaglatcttg 539
Db 547 TGCCTGTGTGTGTGTCCACACT-GCACGCAAGATCACACAGGGGCTTGCACAAAGTATCTTG 605
Qy 540 ctgcagagagaacgcgccttgcacatcttggagaaagcgcttcacaccttgcgtgtgt 599
Db 606 CTGCAGAGAGAGAGCGCGCTGGG-ATCTTGGGAAAGCCGCTTCACCC-TTGCCTGTGTG 662
Qy 600 gcccccac 606
Db 663 GCCCCAA 669

RESULT 15
BE747163 992 bp mRNA EST 15-SEP-2000
LOCUS BE747163
DEFINITION 601577254F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838237 5',
mRNA sequence.
ACCESSION BE747163
VERSION BE747163.1 GI:10161155
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 992)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LUCM524 row: 1 column: 14
High quality sequence stop: 781.
Location/Qualifiers
1..992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3838237"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 225 a 287 c 313 g 167 t
ORIGIN

Query Match 50.5%; Score 505.6; DB 139; Length 992;
Best Local Similarity 97.5%; Pred. No. 96-128;
Matches 546; Conservative 0; Mismatches 9; Indels 5; Gaps 3;
Qy 1 tgcagacacaaagatcaaacacagctcaactcaatccacccggacatcttccccctgt 60
Db 148 TGCACACCCACAGATTCACAAACCCAGCTCAACCTTCATCCACCCGAGCATCTTCCCGCTGT 207
Qy 61. caccagttccgctgtgaagaagagggcccccacccctcaagtgtgcccattgaagtga 120
Db 208 CACCAATTTCCCGTGTGAAGAAGAGGGCCCCACCCCTCACTGTGCCATGTGATGAGGTGA 267
Qy 121 atgcccctccaaagtaccagctccgtccaggaagaggtgcaagaggaatgcatatcac 180
Db 268 ATGCTCTCTCAATACAGATCCGTCCTCCAGAGAGAGTGGCAGAGGATGCCATTATAC 327
Qy 181 ttgcattcttggaggaattatagttgagcgctgtcagcttcccaacttccagacagagcgt 240
Db 328 TTGCATTCCTGAGGATTCATTAGTTAGGGGCTGCAAGCTTCCCACTTCACGACGAGCGT 387
Qy 241 gcaagagtaacagagagagtgctgcagagacgcccagcccaagacagagaaagaagta 300
Db 388 GCAGGAGTACAGAGAGAGAGTGCAGAGACGG-CCAGCCCGCAGCAGAGAGAGAGAGTCA 446
Qy 301 cccagaatcatcttccttggagacaggtctgcacatcccgatgaagattcgaatgtcag 360
Db 447 CCCAGAAATCATCTTCTTGGAAAGGGTCTGCCATCCGATTCGAAATGTGACG 506
Qy 361 tggcaacttgcacataaagcccgacacgctctgtctacttgaactgtgtgagggcac 420
Db 507 TGCCACACTTGTGCACATATAGCCCGACACGTCTTGTGCTACTGAGACTGTGGTGAAGG 566
Qy 421 atttggagacgtgtgcgcgtacattacgagagaccaggtgagacaggtcttggacacctgac 480
Db 567 ATTTGGGAGAGTGTGCCGTCATTACGAGAGCCAGGTGGAGAGGGTCTTGGGCAACCTTGCC 626
Qy 481 tgcgtgtgtgtgtgtccacacttgcacgacagatcaccaacagggcttgcgaagatcttgc 540
Db 627 TGCCTGTGTGTGTGTCCACCT-GCACGCAAGT--CACACACGGCTGCAAGTATCTTGC 682
Qy 541 tgcagagagaacgcgccttgc 560
Db 683 TGCAGAGAGAGAGCGCGCTGG 702

Search completed: October 28, 2001, 21:23:18
Job time: 6051 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 21:48:07 ; Search time 1394.01 Seconds
(without alignments)
11106.963 Million cell updates/sec

Title: SEQINS_COPY_1140_2140
Perfect score: 1001
Sequence: 1 tcgcagccacacagatctcaaa.....aagccatcagcgtggtgagatg 1001

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 8

Total number of hits satisfying chosen parameters: 2254642

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_ba1:*
 - 2: gb_ba2:*
 - 3: gb_ba3:*
 - 4: gb_in1:*
 - 5: gb_in2:*
 - 6: gb_in3:*
 - 7: gb_om:*
 - 8: gb_ov:*
 - 9: gb_pat1:*
 - 10: gb_pat2:*
 - 11: gb_ph:*
 - 12: gb_pl1:*
 - 13: gb_pl2:*
 - 14: gb_pl3:*
 - 15: gb_pl4:*
 - 16: em_bal:*
 - 17: em_ba2:*
 - 18: em_fun:*
 - 19: em_htgo_hum:*
 - 20: em_htgo_inv:*
 - 21: em_htgo_rod:*
 - 22: em_htg_hum1:*
 - 23: em_htg_hum2:*
 - 24: em_htg_hum3:*
 - 25: em_htg_hum4:*
 - 26: em_htg_hum5:*
 - 27: em_htg_hum6:*
 - 28: em_htg_hum7:*
 - 29: em_htg_hum8:*
 - 30: em_htg_inv1:*
 - 31: em_htg_inv2:*
 - 32: em_htg_other:*
 - 33: em_htg_rod:*
 - 34: em_hum1:*
 - 35: em_hum2:*
 - 36: em_hum3:*
 - 37: em_hum4:*
 - 38: em_hum5:*
 - 39: em_hum6:*
 - 40: em_hum7:*
 - 41: em_in:*
 - 42: em_om:*
 - 43: em_or:*

Insertion of
seq in NO 1
oligo search of
1000 residues
packeting
insertion
etc.
at least
8mers

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	880	87.9	2908	89	AF304370 Homo sapi
2	829	82.8	2997	91	BC001939 Homo sapi
3	829	82.8	3006	91	BC004158 Homo sapi
4	750	74.9	2976	89	AK001392 Homo sapi
5	371	37.1	2908	89	AF308698 Pan trogl
6	343	34.3	2893	89	AF308694 Gorilla g
7	124	12.4	118788	85	AC005277 Homo sapi
8	123	12.3	740	89	AF304371S2 Homo sapi

9	39	2712	94	AF308696	Mus muscu
10	35	34593	94	AF348157	Mus muscu
11	23	105490	89	AF165142	Mus muscu
12	23	190705	65	AC019331	Human sapi
13	21	64151	90	AL139399	Human sapi
14	20	1176	8	AB046207	Human sapi
15	20	1573	14	AF023796	Anguilla
16	20	4918	12	AF047031	Arabidops
17	20	6718	85	AB014516	Human sapi
18	20	46275	85	AC003107	Human sapi
19	20	79590	12	AB007727	Human sapi
20	20	84129	12	AB013392	Arabidops
21	20	103638	13	AF14F18	Arabidops
22	20	117338	92	HS173D1	Human sapi
23	20	148508	71	AC027795	Human sapi
24	20	149180	63	AC015525	Human sapi
25	20	150683	74	AC069420	Human sapi
26	20	165493	65	AC018876	Human sapi
27	20	168997	66	AC021927	Human sapi
28	20	177868	69	AC025191	Human sapi
29	20	183412	83	AP001899	Human sapi
30	20	183855	61	AC009591	Human sapi
31	20	184855	67	AC022446	Human sapi
32	20	192059	77	AC090095	Human sapi
33	20	195383	62	AC011804	Human sapi
34	20	197788	80	AL358492	Human sapi
35	20	199812	65	AC018443	Human sapi
36	20	222876	73	AC068379	Human sapi
37	20	222876	73	AC068379	Human sapi
38	20	235141	85	AC004615	Human sapi
39	20	2051	94	AF178954	Mus muscu
40	19	2118	95	RNU87627	Rattus norv
41	19	5841	13	AF136600	Caenorhabdi
42	19	42545	6	CELC32B5	Caenorhabdi
43	19	43390	86	AC007191	Human sapi
44	19	47404	91	AP000333	Human sapi
45	19	61435	78	AC091020	Human sapi

ALIGNMENTS

RESULT	1	2908 bp	mrna	PRI	23-FEB-2001
LOCUS	AF304370				
DEFINITION	Homo sapiens putative prostate cancer susceptibility protein				
ACCESSION	AF304370				
VERSION	AF304370.1				
KEYWORDS	GI:10880932				
SOURCE	human.				
ORGANISM	Homo sapiens				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 2908)
Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M., Ghafiri,S., Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhansen,S., Kommens,J. and Cannon-Albright,L.A.
A candidate prostate cancer susceptibility gene at chromosome 17p
Nat. Genet. 27 (2), 172-180 (2001)

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghafiri,S., Gupta,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhansen,S., Kommens,J. and Cannon-Albright,L.A.
A candidate prostate cancer susceptibility gene at chromosome 17p
Nat. Genet. 27 (2), 172-180 (2001)

Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhansen,S., Kommens,J. and Cannon-Albright,L.A.
Direct Submission
Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt Lake City, UT 84108, USA
lake city, ut 84108, usa
location/Qualifiers
1..2908
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17p"
/chromosome="17"
1..2481
/codon_start=1
/product="putative prostate cancer susceptibility protein
hpc2/elac2"
/translation="MMLCSILRSAGRTMSQGRITISQAPARRPRKDPRLHRTRE
/db_xref="GI:10880933"
/protein_id="AAG24441.1"

CDS

1
10gcagccacaaagattcaaacccagctcaaccctcaccacccgagcatctccctcgt 60
1140 tgcgacccacaaagattcaaacccagctcaaccctcaccacccgagcatctccctcgt 1199
61 caaccgattcgcgtgtaagaagagagccacccctcagtggtccatggtcaggtgta 120
1200 caccagattcgcgtgtaagaagagagccacccctcagtggtccatggtcaggtgta 1259
121 atgctcctcctcaagtagcagctcgcgcacagagagagagagagagagagagagag 180
1260 atgctcctcctcaagtagcagctcgcgcacagagagagagagagagagagagagag 1319
181 ttgcaatcctgaggaatcatagttgagagcgtcgaagcttcccaacttccagcagagct 240
1320 ttgcaatcctgaggaatcatagttgagagcgtcgaagcttcccaacttccagcagagct 1379
241 gcaagagtagcagagagagtagcagagagagagagagagagagagagagagagagag 300
1380 gcagagtagcagagagagtagcagagagagagagagagagagagagagagagagag 1439
301 cccagagagtagcagagagagtagcagagagagagagagagagagagagagagagagag 360
1440 cccagagagtagcagagagagtagcagagagagagagagagagagagagagagagagag 1499
361 tgcgacccacaaagattcaaacccagctcaaccctcaccacccgagcatctccctcgt 420
1500 tgcgacccacaaagattcaaacccagctcaaccctcaccacccgagcatctccctcgt 1559
421 attggcagcgtgtagcgtcattacagagacagagtagcagagagtagcagagagtagc 480
1560 attggcagcgtgtagcgtcattacagagacagagtagcagagagtagcagagagtagc 1619

BASE COUNT 702 a 793 c 826 g 587 t
ORIGIN

Query Match 87.9%; Score 880; DB 89; Length 2908;
Best Local Similarity 99.9%; Pred. No. 0; Mismatches 0; Indels 1; Gaps 1;
Matches 1000; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

|||||
Db 1620 GTTTGGGCGAGCTGTGCGCTCATTTACGAGACACAGGTGAGACAGGCTCTGGGACACCTGGC 1679
OY 481 tgcctgtgttgcctccaccctgcaagcagatcacacacgggcttgccaagtatctgc 540
Db 1680 tgcctgtgttgcctccaccctgcaagcagatcacacacgggcttgccaagtatctgc 1738
OY 541 tgcagagagaacgcgccttgcatcttgggaagccgctcacaccttgcctgtgtgtgc 600
Db 1739 TCAGAGAGAACGGCGCTTGGCATCTTTGGGAAGCGGCTTCACCTTGGTGCTTG 1798
OY 601 cccccaacagctcaaacgcttgctcagcagtlaccacaacagtgccagaggtctgc 660
Db 1799 CCCCCAACAGCTCAAAAGCTGCTCCACAGTACCAACAGTCCAGAGAGTCTGCG 1858
OY 661 accaactcagatgtatcttcgcaaatgctcaggaagggctgagatctcagtcctgc 720
Db 1859 ACCACATCATGATGATCTCTGCGCAAAATGCTTCAGGAAGGGCTGAGATCTCCAGTCTG 1918
OY 721 cagtgaaagatgatcagtcgctgtgtgcgaacatgtatcttgggaagattcgaacct 780
Db 1919 CAGTGAAGAGATTGATGATCTGCTGCTTCGGAACATGATTTGGAGAGTTCAACACT 1978
OY 781 gctcgtgctgcgacgtgcaagcaltgcgttgcctgtgcctgtgcacacctgtgcctga 840
Db 1979 GTCTGCTGCGGCACTGCAAGCATGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2038
OY 841 aagtgctcattccgaggggacacatgctcgtgcagagctgtgtcggatggggaagatg 900
Db 2039 AAGTGTCTATTTCGGGGGACACCATGCTCTGCGAGGCTGTGCTGCGATGGGGAAGATG 2098
OY 901 ccaacctcctgtatcaggaagccacctggaagatgcttgggaaggaagcagtgga 960
Db 2099 CCACCTCTCTGATFACATGAGGACACCTTGGAAGATGTTTGAAGAGGAGCAAGTGGAA 2158
OY 961 agacaacagcaaacgctcccaagccatcagcgtgggagatg 1001
Db 2159 AGACACACAGCACAAGCTCCCAAGCATCAGCGTGGGATG 2199
RESULT 3
BC004158 3006 bp mRNA PRI 16-MAR-2001
LOCUS Homo sapiens, putative prostate cancer susceptibility protein,
DEFINITION clone MGC:2441, mRNA, complete cds.
ACCESSION BC004158
VERSION BC004158.1 GI:13278770
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 3006)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (TLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadanes@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Ketteman and Anuradha Madan
Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/TLNL at: <http://Image.llnl.gov>
Series: IRAL Plate: 2 Row: C Column: 13.
FEATURES
source
1..3006
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:2441 IMAGE:2820640"
/tissue="Lung, small cell carcinoma"
/clone_lib="NIH-MGC-7"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
70..2550
/product="putative prostate cancer susceptibility protein"
/protein_id="AAH04158.1"
/db_xref="GI:13278771"
translation="MMAALCSILRSAGRTMSOGRTISQAPARERPRKDLRLHRTRE
KRPSGCGSGGNTVYLQVNAAGSRDGSALYVSEFNXYLFNCGEVOQLMOEHLKY
ARLDNIFLTRHMSNVGSLSGMILTLKTEGLPKVLSGPOLKYLEAIKIFSGPLK
IELAVRPHSAPVEDEMTVYQIPHSRORRKHOPMSPRPLSRSPERSDSEIN
ENEHLRPHGVSQRRGVRDSSLYVAFICLHKRGNFLYIKAKEMGLPVGTALAPITA
AVDGRSLTHGRREILABELCTPPDPCAAVYVCPDPSF10PICENATFORYGKAD
APVALVHMAPASVLDVSRVQOMEREPDQHLVLENCASVHNLNLSHK1QYDNL1
HPDIFPLTISFRCKREKGTLSVPWVGCECLKYQLRPRERQORDAIITCNPEETVLA
LQLPNFQOVSQYERRSADGPAPAEKRSQYEI1FELGSAIPKIRVSAITVNI5P
DTSLIDGEGTFGOLCRHVGDOVDVRYAAVFAVSLHLDHHTGSLPILLOERLAL
ASLGRPLRLVVAAPNOLKANLQOYHNCQCVLHLSIPKICQBEAEISSPAVERL
ISLRLTDEEFQCLVRHCKHAFGCALVHISGMKVYSQDTPCEALVHMGDNL1
LIHARTJEDGLEEBAVEKTHSTTSQALSVGRNMAEF1MLNHFSQRAKAPLFE5PNS
EKVGVAFFDMKVFGEFPTMPKLI1PPLKALFAGDIEEMEERERREILROYAALLSRE
LAGGLEDPEPOKRAHTEEPQAKVYRAO"
BASE COUNT 742 a 808 c 858 g 598 t
ORIGIN
Query Match 82.8%; Score 829; DB 91; Length 3006;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 999; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
OY 1 tgcagcacaagattcaaacccagctcaacctcaaccgccgagatcttccctcgt 60
Db 1209 TCGCAGCCACAAGATTCAAAACCCACTCATCCACCCGCGACATCTTCCCTGCT 1268
OY 61 caccagtttcgcttaagaagaagggccaccctcagtggtgccatggttcagagtga 120
Db 1269 CACCAAGTTTCGCTGTAGAAGAGGAGGCCCACTCATGTGTGCCATGTGTCAAGGTGA 1328
OY 121 atgcctcctcaagtaaccagctcgcgccagagagagtgagcagaagatgcatatlac 180
Db 1329 ATGCCTCTCAAGTACAGACTCCGTCAGAGGAGAGTGCGAGAGAGTGCATATTATAC 1388
OY 181 ttgcaatccttgaggaattcattagttagagcgctgcagcttcccaacttcagcagagct 240.
Db 1389 TTGCAATCTCTGAGGAATTCAATGATTGAGGCGCTGCAGCTTCCCAACTTCCAGCAAGAGCT 1448
OY 241 ggaagaatgacaggaagagtgagcagaagcggccacccacagagaagaagaagacata 300
Db 1449 GCAGGAGTACAGAGGAGGCGCCAGAGAGGCCACCCACGAGAGGAAAGATCAGTA 1508
OY 301 cccaagaatcattcctcttgagacagggctgcacatcccgatgaagattcgaaatgcag 360
Db 1509 CCGCAAAATCATCTTCTTTGGAACAGGGGTCTGCCATCCCGATGAAGATTCCGAAATGTCAG 1568
OY 361 tgcacactgtcaacataaagcccgacaagctctctgctactgtagctgtgtgagggcac 420
Db 1569 TGCCACACTTGTCAACATAAAGCCCGACACAGTCTGCTAGCGACTGTGGAGAGGAC 1628
OY 421 atttggaagctgttcgctcatttagagagaacaggttgacaaggtccctgggcaacctggc 480
Db 1629 GTTTGGGAGCTGTCCGCTTACGAGACCAAGGTGAGAGGCTCTGGGACCTCTGGC 1688
OY 481 tgcctgttgcctccaccctgcagcagatcaccaacacagggcttgccaagtatctgc 540

```
Db 1689 TGCTGTGTTGTGTTCCACCT-GCAGCAGATACACACAGGGCTTGCCAGATATCTTG 1747
Qy 541 tgcagagagaagcgcccttggcatcttgggaagccgcttcaaccttgcgtggttg 600
Db 1748 TCGAGAGAGACGCGCTTGGCATCTTTGGGAAGCCGCTTCACTTGTGTTG 1807
Qy 601 cccccaacagctcaagctgtctcagcagctacacacacagctccagaggtcttcg 660
Db 1808 CCCCCAACCACTCAAGCTTGCTCCACAGTACCAACAGGTCCAGAGAGTCTTG 1867
Qy 661 accacacagatgatttcttcccaaatgcttcaagaagggtcagatctcagctcgt 720
Db 1868 ACCACATCATGATGATCTCTGCCAAATGCTTCAGAGAGGGCTGAGATCTCCAGTCTCG 1927
Qy 721 cagtggaagattgacatgctgctgtctgcgaacatgtgatttgaagagttcagacct 780
Db 1928 CAGTGGAAAGATTGATGACATGCTGTGCGAACAATGTATTTGGAAGATTTCAGACCT 1987
Qy 781 gctcgtgctggcactgcagcagatcgctgtgctgtgctgtgcacacctctgcgtga 840
Db 1988 GCTCTGCTGGCGCACTGCAAGCATGCTTTGGCTGTGCTGCTGTCACACCTCTGCTGGA 2047
Qy 841 aagtgatctatccggggagacacatcgcctgcgaggctctgtcgcgagatgggaagatg 900
Db 2048 AAGTGTCTATTCGCGGGAGACACATGCTCCTCGAGGCTCTGCTGCGGATGGGAAAGATG 2107
Qy 901 ccaacctctgatacagaagaacacccctggaagatggttggagaagaagcagtggaaga 960
Db 2108 CCACCTCTCTGATACATGGAAGCCACCTCGAAGATGTTGGAAGAGAAAGCAGTGGAAA 2167
Qy 961 agagacagagcacaagctcccaagcagcagtggtggtg 1001
Db 2168 AGACACACAGACACACGCTCCCAAGCATCAGCTGGGAGT 2208
```

```
source 1.. 2976
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone="NT2RP2000985"
/clone_1ib="NT2RP2"
/note="cloning vector: pME18SFL3: mRNA from NT2 neuronal precursor cells after 2 weeks retinoic acid (RA) induction."
75.. 2639
/cds="unnamed protein product"
/code_start=1
/protein_id="BAA91666.1"
/db_xref="GI:7022622"
/translation="MVALCSLRSASGRMSOGRTISQAPARREPRKDPRLHLRTRE
KRPSGCSGPNVTVLYQVVAAGSRDSGALVYSEFNRLPFCGCGVORLWQENLKY
ARLDNIPLTRMIMSNVGLSGMILTKGLEPKCVLSGPOLKYLEAIKIFSGPLK
IELARPHSAPSEYEDEMTVVOIPIHSRORRKHQPMOSPERPLSRSPERSDES
ENRPLPHGVSGRRGVRDSLVAVATFCKLHLKRGNFVLKAKEMGLPVGTAAIARIT
AVKDKSTHBSRELIAELCTPPRPGAAYVVECPDPSFTPTCENNTFOYQKAD
APVALVMAAPASVLDSTRYQMMERFGDTQHLVNECASVHNLRSKIDTQNLTI
HPDIPPLTSEFRCKEGPTLSVPMVQECLELKYQLPRRMWDALITGNPEEIVLAL
OLPNFOQSOYEYRERSADGPAPAEKRSQPELIFLGTSALIPKIRNYSATLVNISPD
TSLLDGCGFTGOLCRHGDVDVRLGTIAVFSHLADHHTGLPSILORBALA
SIGKPLHLVYVAPNOLKAMLDQVYNOCQEVVHHTSMTPAKTOGSAETSPAVERLI
SGLRTCDLEERTCLVHRCKHAFECALVHTSGMKVYVSGDTMPCPALRMKQDITL
IHEATLEGLDEEAVERKTHSTSOAISVMGRNNAEFLINLNEFSRYAVVPLEPNSE
KVGVAFDHMKVCLGDFPMPKLIPLKALFAGDIEEMERREKRLROYRALLSREL
AGGLDGSISRSRGPORSHRPRSPSEDLGPELRILCVFCPTBAVPSALLAGRS"
BASE COUNT 712 a 807 c 856 g 601 t
ORIGIN
```

Query Match 74.9%; Score 750; DB 89; Length 2976;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 920; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
RESULT 4
AK001392
LOCUS AK001392 2976 bp mRNA PRI 22-FEB-2000
DEFINITION Homo sapiens CDNA FLJ10530 fls, clone NT2RP2000985.
ACCESSION AK001392
VERSION AK001392.1 GI:7022621
KEYWORDS oligo capping; fls (full insert sequence);
SOURCE Homo sapiens teratocarcinoma cell line:NT2 CDNA to mRNA,
clone_1ib:NT2RP2 clone:NT2RP2000985.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Matanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahari,K., Masuno,Y., Ninomiya,K. and Iwayanagi,T.
TITLE NED0 human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2976)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory: 152-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NED0 human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing; Research Association for Biotechnology: cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
```

```
Qy 80 aagagaggcccaacctcagtggtcccatggttcaggtgtaatgctctcctaagtcag 139
Db 1290 AAGGAGGGCCCCACCTCAGTGTGCCATGCTTCAGGGTGAATGCTCTCCAGTACCAAG 1349
Qy 140 ctccgtcccaaggagagatgagcaggagatgcatattacttgcaactcttgagaattc 199
Db 1350 CTCCTGCCAGGAGGAGGAGGAGGAGGATGCCATTATTACTTGCAATCTTGAGAAATTC 1409
Qy 200 atagttgagcgctgtcagcttcccaactccagcagagcgtgcaggaagtcacaggaagat 259
Db 1410 ATAGTTGAGGCGCTGACAGCTTCCCACTTCCACACAGAGGCTGACAGAGATNACAGAGAGAT 1469
Qy 260 ggcgaggaagcccaagcccaagcagagaagaagaagtcagtaaccagaatcatcttcctt 319
Db 1470 GCCCAGAGCGCCACCCACCCAGCAGAGAAAGAACAGTACAGCACCAGAAATCATCTTCCCTT 1529
Qy 320 ggaagagggtctgtccatcccatgaagaattcgaaatgtaagtcgacacttgcacaata 379
Db 1530 GGAACAGGGGTCTCCATCCCGATGAAGATTGCAAAATGTAGTCCACACTTGTCACACATA 1589
Qy 380 agccccgaacgctctctgtactgtgactgtgtgtgagagcatttggagcagctgtgcgt 439
Db 1590 AGCCCGGACACGCTCTGTGACTGAGACTGTGTGAGAGGACATTTGGGACGCTGCCGT 1649
Qy 440 cattacgagacagatgtgacaggtctcgtggcaccctgtcgtgtgttgtgtccac 499
Db 1650 CATTACGAGACACAGGTGACAGGCTCTGGGACACCTGTGCTGTGTTGTGCTCCAC 1709
Qy 500 ctgcgacgaagatcacacaacaggcgttccaagtatctgtgtcagaagaagcagcctt 559
Db 1710 CT-GCAGCAGATCACACACGGGCTTGCCAAATATCTTGCTGCAAGAGAAAGCCCTTT 1768
Qy 560 ggcattcttggaagacgcttcaaccttgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 619
Db 1769 GGCATCTTTGGAAACCGCTTACCTTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1828
```

Qy	620	ctgtctccagcagtlaccacaacagtcgcaaggaggttcctgtcacccacagatgatgtcc	679
Db	1829	CTGGCTCCAGAGTACTAACAACAGTGCAGAGAGTCCGACACATCAGATGATTTCC	1888
Qy	660	tcgcaaatgctctcaggaagggtctgagatctccagtcctcagtggaagaatgatcaag	739
Db	1889	TGCCAAATGCCCTTCAGGAGAGGGCTGAGATCTCCAGTCCCTGAGTGAAGAATGTGATCAG	1948
Qy	740	ttgctgtttgcaaatctgatttggaagaagtttcaaaccttctcgtgtggcgactgcaa	799
Db	1949	TTGCTGTGTCGAACATGATGATTTGGAAGATTTTCAGACTCTGTCTGGTCCGCACTGCAA	2008
Qy	800	gcatacgtttgtgctgtgctgtgtcacacacctgtgcttggaagaatggtctatctccgagga	859
Db	2009	GCATCGCTTTGGCTGTGCGCTGTGTCACACACTCTGGCTGGAAGAAGTGTCTATTCGGGGGA	2068
Qy	860	caccatgctctcgagagctctgtgtccgagatggggaagaatgcccacctctgatatcatga	919
Db	2069	CACCATGCCCTCGAGGCTCTGTGTCGGATGGGGAAGATGCCACCTCTGTATCATATGA	2128
Qy	920	agccacccctggaagaatggttttggaagaagaagcagtggaagaagacacacagcaagctc	979
Db	2129	AGCCACCTTGGAAGATGCTTTGGAAGAGGAACAGTGGAAAAGACACACAGCAACAGCTC	2188
Qy	980	ccaagccatcaacgctgaggaatg 1001	
Db	2189	CCAAGCCATCAACGCTGAGGATG 2210	
RESULT	5		
AF308698	2908 bp	mrna	PRI 27-FEB-2001
LOCUS	Pan troglodytes ELAC2 mRNA, complete cds.		
DEFINITION	AF308698		
ACCESSION	AF308698.1	GI:10946496	
VERSION			
KEYWORDS			
SOURCE	chimpanzee.		
ORGANISM	Pan troglodytes		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan. 1 (bases 1 to 2908)		
	Tavtigian, S.V., Simard, J., Teng, D.H.F., Abtin, V., Baumgard, M., Beck, A., Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaffari, S., Gupta, J.S., Hu, R., Iliev, D., Janacki, T., Kort, E.N., Laitty, K.E., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R., Snyder, S.C., Swedlund, B., Swensen, J., Thomas, A., Tranchant, M., Woodland, A.M., Labrie, F., Skolnick, M.H., Neuhausen, S., Rommens, J. and Cannon-Albright, L.A.		
	A candidate prostate cancer susceptibility gene at chromosome 17p Nat. Genet. 27 (2), 172-180 (2001)		
TITLE	JOURNAL		
REFERENCE	PUBMED		
AUTHORS	2 (bases 1 to 2908)		
	Tavtigian, S.V., Simard, J., Teng, D.H.F., Baumgard, M., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaffari, S., Gupta, J.S., Hu, R., Iliev, D., Janacki, T., Kort, E.N., Laitty, K., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R., Snyder, S.C., Swedlund, B., Swensen, J., Thomas, A., Tranchant, M., Woodland, A.-M., Labrie, F., Skolnick, M.H., Neuhausen, S., Rommens, J. and Cannon-Albright, L.A.		
	Pan troglodytes ortholog of human HPC2/ELAC2 unpublished 3 (bases 1 to 2908)		
TITLE	JOURNAL		
REFERENCE			
AUTHORS	Tavtigian, S.V., Simard, J., Teng, D.H.F., Baumgard, M., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaffari, S., Gupta, J.S., Hu, R., Iliev, D., Janacki, T., Kort, E.N., Laitty, K., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R., Snyder, S.C., Swedlund, B., Swensen, J., Thomas, A., Tranchant, M., Woodland, A.-M., Labrie, F., Skolnick, M.H.,		

TITLE	Neuhausen, S., Rommens, J. and Cannon-Albright, L.A.		
JOURNAL	Direct Submission		
FEATURES	Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt Lake City, UT 84103, USA		
source	Location/Qualifiers		
	1..2908		
	/organism="Pan troglodytes"		
	/db_xref="taxon:9598"		
CDS	1..2481		
	/codon_start=1		
	/product="ELAC2"		
	/protein_id="AAG24920.1"		
	/db_xref="GI:10946497"		
	/translation="MMAICSLRSAAGRFMSQGRFISQAPRARERPRKDPRLRLTRRE KRSGSGSGGNFVYLQVVAAGSDGSAALYFSEFRLFNCGEGLQRLMOEHLKY ARLDNITLRTMHSNNGSLGMLITLKECTLPKCVISGSPOLKEYLEAIFISGPKG TELAVRPHSAPEYDEITMYQIPIHSEQRGRKQPMQSPERPLSLRSPERSSEEN ENEPHLRHGVQSQRGVRDSSLVAVFICKLHKRGNFLVLRKENGFLVGTALAPITA AVKDKSITHEGREILAELELCTPPDPAARVAVVCEPDEFTOPICENATQROKGD APVALVVMHAPESYLVDSRYQOMMEREGPDTQVLNENCASVHNLRSKIQIOLNLI HPDIFPLTSPCKEKGPTLSVPMVOGECILKQLRBRREMRDAITTCNPEEFIIA LQLPNPOSVQEVYRRSADQDPAPAKERSQYPELIFGTGSAIPMKIRNVSATLVNISB DTSLLDCEGCTRGQLCRHGYDVRVLTGLAIVFVSHLADHTGLNLTLQERALL ASLQKPHPLVVAPNQKAWLQDYHMOCEVLHISMIPAKQLQEGAEISSPAVERT ISSLRTCDLEEFQCLVRHCKAFGCALVTSQWKVYSGDTWPCALVBMGDATL LHEATVLEDGLIEEAVERHTSTTSQAISVGMRAAEFIMLNHFSQRAKVPILSPFN EKVGAFDHMKVCGDFATMPRLIPPKALFACDIEEMERERERELROYRAALLSRE LAGSLDEGEPOQRKRAHTEEPQAKVVRQ"		
BASE COUNT	712 a 788 c 819 g 589 t		
ORIGIN			
Query Match	37.1%; Score 371; DB 89; Length 2908;		
Best Local Similarity	99.8%; Pred. No. 2.9e-205;		
Matches	421; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy	580	ttcaccccttgctgtgtgtgtccccaacagctcaaaagctgtgctcagcagttaccaca	639
Db	1718	TTCAACCTTTGCTGTGTGTGTGCGCCCAACAGCTCAAGCCTGCGTCCAGAGTACACA	1777
Qy	640	accagtgccaaggaggtctctgtcacccacatcatgatgtatcttcgtcccaaatgcttcaaggag	699
Db	1778	ACCACTGCCAGAGAGTCTCTGACCAACATCATATGATTTCTGCCAAATGCTTCAGGAAG	1837
Qy	700	ggcgtgagatctccagctctcagtggaagaatgatgacgttgcgtgttggaacatgtg	759
Db	1838	GCGCTGAGATCTCCAGCTCTGCTGCAAGTGAATGATCATGTTGCTGTGGCAACATGTG	1897
Qy	760	atttggaagaatttcagacctgtctgtgtgcggcactgtgcaagcatgctgtgtgtgtgcgc	819
Db	1898	ATTTGGAAGAGTTTCAGACCTGTCTGTGCGGCACTGCCAAGCATGCTGTTGGCTGTGCGC	1957
Qy	820	tgtgtgcaacctgtgctgtggaagtgtctatctccggggaacacatgcccctgcgagagctc	879
Db	1958	TGCTGCACACCTCTGCTGCGGAAGTGTCTATTCGGGGGACACCATGCTGCGAGGCTC	2017
Qy	880	tgttcgagatgtggaagaatgccacccctcgatatcatgaaagccaccccttggaagaatgtt	939
Db	2018	TGCTCGGATGGGGAATAATGTCACACCTCTGTATCATGAAAGCCACCTGGAGAAGCGTT	2077
Qy	940	tggaaaggaagcagtggaagaagacacacagcaacaagtcctccaaagcatcagcgttgagga	999
Db	2078	TGGAAGAGGAAGCAAGTGGAAAAGACACACAGCAACAGTCCCAAGCCATCAGCGTGGGGA	2137
Qy	1000	tg 1001	
Db	2138	TG 2139	
RESULT	6		
LOCUS	AF308694	2893 bp	mrna PRI 27-FEB-2001
DEFINITION	Gorilla gorilla ELAC2 (ELAC2) mRNA, complete cds.		

ACCESSION AF308694
VERSION AF308694.1 GI:10946488
KEYWORDS
SOURCE
ORGANISM gorilla.
gorilla.gorilla
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Gorilla.
REFERENCE
AUTHORS Tavtigian, S.V., Sismard, J., Teng, D.H.F., Abtlin, V., Baumgard, M., Beck, A., Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaffari, S., Gupte, J.S., Hu, R., Iliev, D., Janacki, T., Kort, E.N., Laitly, K.E., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Schroeder, M., Smith, R., Snyder, S.C., Sweldund, B., Svensen, J., Thomas, A., Tranchant, M., Woodland, A.-M., Labrie, F., Skolnick, M.H., Neuhausen, S., Rommens, J., and Cannon-Albright, L.A.
A candidate prostate cancer susceptibility gene at chromosome 17p11.7585
Nat. Genet. 27 (2), 172-180 (2001)
TITLE
JOURNAL PUBMED
REFERENCE
AUTHORS 2 (bases 1 to 2893)
Tavtigian, S.V., Sismard, J., Teng, D.H.F., Baumgard, M., Beck, A., Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaffari, S., Gupte, J.S., Hu, R., Iliev, D., Janacki, T., Kort, E.N., Laitly, K.E., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R., Snyder, S.C., Sweldund, B., Svensen, J., Thomas, A., Tranchant, M., Woodland, A.-M., Labrie, F., Skolnick, M.H., Neuhausen, S., Rommens, J., and Cannon-Albright, L.A.
Gorilla gorilla ortholog of human HPC2/ELAC2
Unpublished
TITLE
JOURNAL
REFERENCE
AUTHORS 3 (bases 1 to 2893)
Tavtigian, S.V., Sismard, J., Teng, D.H.F., Baumgard, M., Beck, A., Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaffari, S., Gupte, J.S., Hu, R., Iliev, D., Janacki, T., Kort, E.N., Laitly, K.E., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R., Snyder, S.C., Sweldund, B., Svensen, J., Thomas, A., Tranchant, M., Woodland, A.-M., Labrie, F., Skolnick, M.H., Neuhausen, S., Rommens, J., and Cannon-Albright, L.A.
Direct Submission
Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt Lake City, UT 84103, USA
FEATURES
source Location/Qualifiers
1. .2893
/organism="Gorilla gorilla"
/db_xref="taxon:9593"
1. .2893
/gene="ELAC2"
1. .2481
/gene="ELAC2"
/note="similar to Homo sapiens HPC2/ELAC2"
/codon_start=1
/product="ELAC2"
/protein_id="AAC24916.1"
/db_xref="GI:10946489"
/translation="MVALCSLRSAAGRTMSOGRTISQAPARERPRKDLRLHRTREKPGSCGSPNTVLYOYVAGSRDGAALVSEFNRLFPNGGEVLOMDEHKLYVLDNFTLRHMSNVNGLSGMLITLKEGILPKCYLSPGPOLEKYEALIKITSPKIGITLAVRPHAPPEDEETMYVYIPIHSEORRGHOPWSPERPLSLSPERSDSESNENEPHLPVGSORRGVRSLSLVAATCKLHLKGNFLVLAKEMGVPTGTAIAPLIAAVAKDSITHEREILAEELCPDPGAFVVECEPDESIOPICNAPETAFOROGADAVALAIVHMAPELSVDSRYOQMERGFDTOLHLYNENCAVSHNRSKIQTLNLIHDDIPPLTSPCKKEGPTLSVPMWEGCLLYKLRPREMDADITTCNPEEPIVEA IOLPNQOSVOEYRSYODVPAAPKRSOYPELITIGCSGASIPKIRANSATLVNISP DTSLLDCGEGTGGOLCRHYGQVDRVLTGLAAYVSHLHADHHTGLTILQREDAI ASLGRPLRLTVVAPSQLKAWLQOYHNQOEVLNHSIMPAACLEGAGTISPAVERL IISLTLTDLGEFOCTLVKCHKAFCAVYHLSGMKVYSGDTPMEALVVRGKDALI LHEATLEDGEEAEVAKHTSTSOAISVGMNMAEFILNHSNOCPEALVYFSPEN EKVGVAFDHMKVCGDFPTMPKLIPLIKLAFAGDIEEMERREKRLRQVRAALLSGE LAGLEDEGRPOCKRAHTEPRQAKYKRAO"

BASE COUNT 705 a 787 c 815 g 586 t
ORIGIN
Query Match 34.3%; Score 343; DB 89; Length 2893;
Best Local Similarity 99.7%; Pred. No. 7, 2e-189;
Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 608 ccagctcaaaagcttgctccagcagatccacacacccagtcaggaagctctgaccacat 667
|||||
Db 1746 CCAGCTCAAAAGCTTGCTCCAGACGACCAACCAACGATGCGAGAGTCTTGACACAT 1805
QY 668 cagatgattccctgcgaatgcttcaaggaagggtcgaatccagtcctcagtgga 727
|||||
Db 1806 CAGTATGATTCTCTGCCAATGCTTCAAGAAAGGCGGTGACATCTCCAGTCTCATGTGA 1865
QY 728 aagatgattcagtcgtctgttcgaaacatgtgatttggaagagtttcagacctgtctgt 787
|||||
Db 1866 AAGATGATCAGTTGCTGTGTGCAACATGTGATTTGGAAGAGTTTCAGACCTGTCTGTGT 1925
QY 788 ggcgcactcgaagcagatgctgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 847
|||||
Db 1926 GCGGCACTGCACACATGCGTTTGCTGCTGCGCTGCGACACCTGCTGCGAAAGTGTGT 1985
QY 848 ctattccggagacacacatgctgcctgcagagctcgtgtcccgatgggaagaatgccacct 907
|||||
Db 1986 CTATTCCGGGACACCATGCGCTTGCAGGCTCTGTCCGATGGGGAAGATGCCACCT 2045
QY 908 cctgatcacatgaagccacaccttggaagaatgttggaagaagagcagttggaagaagacaca 967
|||||
Db 2046 CCGATGATGATGAGACGACACCTCGAAGATGTGTTGGAAGAGAGCAGTGAAGAAAGACACA 2105
QY 968 cagcacacagctcccaagcaccatcagctggtggagatg 1001
|||||
Db 2106 CAGCACACAGCTCCCAAGCCATGACGCTGGGGATG 2139
RESULT 7
AC005277 118788 bp DNA PRI 23-JUL-1998
LOCUS Homo sapiens chromosome 17, clone hRPK.597_M.12, complete sequence.
DEFINITION AC005277
AC005277
VERSION AC005277.1 GI:3337311
KEYWORDS HPG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Birren, B., Fasman, K., Linton, L., Nusbaum, C., Lander, E., Allen, N., Binkov, J., Baldwin, J., Barna, N., Beckwith, R., Benn, J., Boatin, C., Boutwell, C., Brown, A., Castle, A., Cerny, D., Cooke, P., Depierre, E., Devon, K., Dewar, K., Donelan, L., Etemadi, S., Ferrelita, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Garavito, S., Genshler, S., Gerahty, K., Gilmer, T., Grant, G., Hagos, B., Harris, K., Horton, L., Howland, J.C., Hui, L., Jacotot, L., Kann, L., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Nachman, A., Nair, R., Naylor, J., Nilot, M., O'Connor, T., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Rossello, R., Roy, A., Shyam, R., Stange-Thomann, N., Stilwell, J., Stojanovic, N., Stone, C., Strickland, C., Subramanian, A., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wang, B., Wheeler, J., Wu, Y., Ye, W., Zhao, J. and Zody, M.
Direct Submission
Submitted (10-JUL-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
AUTHORS Birren, B., Fasman, K., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

```

Repeat_region      8039. .8333  
/rpt_family="AluBq"  
9431. .9460  
/rpt_family="AT-rich"  
9662. .9829  
Repeat_region  
/rpt_family="MIR"  
complement(11302. .11728)  
/rpt_family="MER39"  
complement(11824. .11884)  
Repeat_region  
/rpt_family="FTR16C"  
complement(11928. .12108)  
Repeat_region  
/rpt_family="FTR16C"  
12453. .12725  
Repeat_region  
/rpt_family="AluBb"  
13377. .13785  
Repeat_region  
/rpt_family="MSTB"  
14118. .14153  
Repeat_region  
/rpt_family="(CA)n"  
complement(15320. .15620)  
Repeat_region  
/rpt_family="AluDo"  
15683. .15759

```

```
repeat_region complement(15838. .15893)
repeat_region /rpt_family="MER91A"
repeat_region complement(15915. .16080)
```

Location/Qualifiers
1. .118788

1. .118788

```

795. .884
    /rpf_family="MIR"
    complement(932. ,1000)
    /rpf_family="MIR"
    2044. .2076
    /rpf_family="AT_Rich"
    2237. .2713
    /rpf_family="MER34"
    2718. .3477
    /rpf_family="L2"
    complement(3478. ,3821)
    /rpf_family="MLT1A1"
    3822. .4007
    /rpf_family="L2"
    complement(4015. ,4281)
    /rpf_family="MER89"
    complement(4282. ,4321)
    /rpf_family="(TAAAn)"
    complement(4328. ,4608)
    /rpf_family="AluSx"
    complement(4609. ,4895)
    /rpf_family="MER89"
    complement(4896. ,5186)
    /rpf_family="AluSp"
    5219. .5479
    /rpf_family="L2"
    complement(5481. ,5740)
    /rpf_family="MER4"
    complement(5963. ,6275)
    /rpf_family="AluSp"
    complement(6411. ,6494)
    /rpf_family="MER8A"
    6830. .7013
    /rpf_family="MLT1C"
    7014. .7318
    /rpf_family="AluYb8"
    7319. .7581
    /rpf_family="MLT1C"
    7721. .8032
    /rpf_family="AluYb"

```

```

8039 . 8333
/rfp_family="AluSq"
9431 . 9460
/rfp_family="Ar_rich"
9662 . 9829
/rfp_family="MR"
complement(11302 . 11728)
/rfp_family="MER39"
complement(11824 . 11884)
/rfp_family="LRR1cC"
complement(11928 . 12108)
/rfp_family="LRR1cC"
12453 . 12725
/rfp_family="AluJb"
13377 . 13785
/rfp_family="MSTB"
14118 . 14153
/rfp_family="(CA)"
complement(15320 . 15620)
/rfp_family="AluJc"
15683 . 15759
/rfp_family="(CA)"
complement(15838 . 15893)
/rfp_family="MER91A"
complement(15915 . 16080)
/rfp_family="MER91A"
complement(16474 . 16544)
/rfp_family="L2"
17907 . 17920
/rfp_family="LIME"
18271 . 18311
/rfp_family="L2"
complement(18368 . 18529)
/rfp_family="AluJb"
complement(18530 . 18560)
/rfp_family="(CA)"
complement(18561 . 18681)
/rfp_family="AluJb"
18760 . 19065
/rfp_family="AluSg"
complement(19638 . 19896)
/rfp_family="MR"
complement(19971 . 20260)
/rfp_family="LIME3A"
20263 . 20552
/rfp_family="MERS8B"
complement(20553 . 20816)
/rfp_family="LIME3A"
complement(20821 . 21065)
/rfp_family="LIM4"
complement(21071 . 21100)
/rfp_family="(TAA)n"
complement(21101 . 21338)
/rfp_family="AluSx"
complement(21348 . 21598)
/rfp_family="LIM4"
21799 . 22104
/rfp_family="AluJc"
22151 . 22562
/rfp_family="MSTA"
23978 . 24287
/rfp_family="AluSg"
24852 . 25189
/rfp_family="L2"
complement(25697 . 25985)
/rfp_family="MULTIP"
25816 . 25895
/rfp_family="Single-stranded coverage"
complement(25994 . 26099)
/rfp_family="purine-rich"
complement(26122 . 26291)
/rfp_family="MULTIP"
26312 . 26337

```

```

repeat_region      /rpt_family="(CAAAA)n"
26456..26480
repeat_region      /rpt_family="(GGGAA)n"
complement(26576..26781)
repeat_region      /rpt_family="MER6B"
complement(26912..26990)
repeat_region      /rpt_family="MER21B"
26991..27290
repeat_region      /rpt_family="Alusx"
complement(27291..27975)
repeat_region      /rpt_family="MER21B"
28034..28197
repeat_region      /rpt_family="LTR28"
28266..28795
repeat_region      /rpt_family="LTR28"
28843..28916
repeat_region      /rpt_family="LTR28"
28920..29005
repeat_region      /rpt_family="MIR"
29140..29233

Query Match      12.4%; Score 124; DB 85; Length 118788;
Best Local Similarity 100.0%; Pred. No. 1.1e-60;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 agttcagacctgtctgtcgcgcacgcacgcacgtcgttgcgtgtgcgtgtgcaca 828
|||||
Db 101623 AGTTTCAGACCTGTCTGTCTGCGGCACATGCATGCGTTGCTGCGTGTGCGACA 101682
|||||
QY 829 cctcgtcgtgaagtgctgtctatccgcggagacacacgcctcgtcgtcgtcgtcgcgga 888
|||||
Db 101683 CCTCTGCTGTGAAGGTGTCTATTTCCGGGACACCATGCCCTCGAGAGCTCTGTGTCGGA 101742
|||||
QY 889 tggg 892
|||||
Db 101743 TGGG 101746

RESULT 8
LOCUS      AF304371S2      740 bp      DNA
DEFINITION Homo sapiens putative prostate cancer susceptibility protein
ACCESSION AF304369
VERSION AF304369.1 GI:10880929
KEYWORDS   2 of 2
SEGMENTS
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 740)
AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
Dumont,M., Farnham,J.M., Frank,D., Frye,C., Gaffari,S.,
Gupte,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E.,
Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
Syder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
and Cannon-Albright,L.A.
A strong candidate prostate cancer susceptibility gene at
chromosome 17p
Unpublished
JOURNAL    2 (bases 1 to 740)
REFERENCE  Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
AUTHORS   Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
Dumont,M., Farnham,J.M., Frank,D., Frye,C., Gaffari,S.,
Gupte,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E.,
Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
Syder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
and Cannon-Albright,L.A.

```

```

TITLE      and Cannon-Albright,L.A.
JOURNAL    Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
Lake City, UT 84108, USA
FEATURES   Location/Qualifiers
Source     1..740
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="17"
            /map="17p"
            /number="17"
            /product="putative prostate cancer susceptibility protein
            HPC2/ELAC2"
            /protein_id="FAG24440.1"
            /translation="PDTSLLDGECSTFQQLCRHYDDYDRVLGTLAAVFSHLHADI
            HT"
BASE COUNT 154 a 188 c 173 g 225 t
ORIGIN

Query Match      12.3%; Score 123; DB 89; Length 740;
Best Local Similarity 100.0%; Pred. No. 4.1e-60;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 agcccgacacgctctcgtcactggaactggtgagggacacattggcgcgtgtgcgt 439
|||||
Db 299 AGCCCGACACGCTCTGCTACTGACTGTGTCAGGACATTTGGGACGCTGTGCGGT 358
|||||
QY 440 cattacgagacacgagtggtgcacggtccttggtgcacccctgctgtgtgtgtccac 499
|||||
Db 359 CATTACGAGACACGAGTGTGACAGGGTCTGTGGCACCTGTGCTGTGTGTGTCAC 418
|||||
QY 500 ctg 502
|||||
Db 419 CTG 421

RESULT 9
LOCUS      AF308696
DEFINITION Mus musculus ELAC2 mRNA, complete cds.
ACCESSION AF308696
VERSION AF308696.2 GI:11992378
KEYWORDS
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2712)
AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtl,V., Baumgard,M.,
Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
Gaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N.,
Lally,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
A candidate prostate cancer susceptibility gene at chromosome 17p
Net. Genet. 27 (2), 172-180 (2001)
JOURNAL    2 (bases 1 to 2712)
REFERENCE  Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
AUTHORS   Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
Dumont,M., Farnham,J.M., Frank,D., Frye,C., Gaffari,S.,
Gupte,J.S., Hu,R., Iliev,D., Janacki,T., Kort,E.N., Lally,K.E.,

```

Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.,
and Cannon-Albright,L.A.
Mouse ortholog of human HPC2/ELAC2

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
Lake City, UT 84103, USA
4 (bases 1 to 2712)
Tavligian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
Gupte,J.S., Hu,R., Iliev,D., Janekil,T., Kort,E.N., Laity,K.E.,
Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.,
and Cannon-Albright,L.A.
Direct Submission
Submitted (26-DEC-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
Lake City, UT 84103, USA
Sequence update by submitter
On Dec 26, 2000 this sequence version replaced gi:10946492.

TITLE
JOURNAL
REMARK
COMMENT
FEATURES
SOURCE

CDS

Location/Qualifiers
1..2712
/organism="Mus musculus"
/db_xref="taxon:10090"
/dev_stage="fetus"
15..2510
/note="similar to Homo sapiens HPC2/ELAC2"
/codon_start=1
/product="ELAC2"
/protein_id="AAG24918.2"
/db_xref="GI:11992379"
/translation="MMAIRSLIRPGLRTMSGSAARRPRPKDPIRLHRTREKRCPP
GCPNTVYLQVVAAGGRDGAALYFSEYNRIYFNGEGVORLMOEHLKLVADNIFL
TRMHSNVGGLGMLTLKETGLPCVLSGPQLKYLEAIRKIFSGPKGLELAVRPH
SAPEKDETMVYQVPIHSEKRCQKQSPRTSPNRLSPKQSSDSCSAEGQCPE
DSAGANKRANGRDPSLVAAFCVCKLHKKGNFLVAKELGLPVGTAIAPIAAVKD
GKSTYEGREIAAEELCTPPDGLVIFVECDDEEFLIPICNDREKRYOAEADPA
LVYHIAPSVILDSRYOQMMEREGDTQHLINENCPVHNRSKIKTOGSLIHPDI
EPQUTSTSKBEGSTLSVPIYRGECLKYQLRKREKWDRTDPDNDTEFLAEALELP
SFQSEVEYRNKNQENPAAPKRSQYPRIVELGTSALPMKIRNVSITLVNISPVS
LLDCEGTFGQICRHYGOQIDRVLSLTAVESVSHADHTGLNITLLQREHALASIG
KPPPLVAVAPTOLRAMLOQYHNHCOEILHNHSMIPAKCLOGAENVNTTERLISIL
LETQDLFEFOFICVIRHCKAFHGCALVHSSGKVVYSGDPMCEALVOMGKATLLIIE
ATLEDGLEEPAVERKTHSTTSOAINGNMNGEITLNNFSORYATIRLPSPDRNKVC
IATDHMKVCFGDFPTVPKLIPLPLKLFAGDIEEMVEREKREKLRLVALLLQOQDSP
EDREPOQRAITDEPHSPQSKSESVANTLGARV"

BASE COUNT
ORIGIN

Query Match 3.9%; Score 39; DB 94; Length 2712;
Best Local Similarity 100.0%; Pred. No. 6e-11;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 887 gatgggaaagatgcacacctctgatacatatgaagccac 925
DB 2027 GATGGGAAAGATGCCACCTCTGTATCATGAGCCAC 2065

RESULT 10
AF348157
LOCUS
DEFINITION
MUS musculus putative prostate cancer susceptibility protein
(Elac2) gene, complete cds, alternatively spliced.
AF348157
AF348157.1 GI:13540341

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janekil,T., Kort,E.N.,
Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
Mouse Elac2-containing genomic DNA

TITLE
JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 34593)
Frank,D.C., Swedlund,B., Dumont,M., Tavligian,S.V., Simard,J.,
Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janekil,T., Kort,E.N.,
Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
Direct Submission
Submitted (09-FEB-2001) Myriad Genetics, Inc., 320 Wakara Way, Salt
Lake City, UT 84108, USA
Location/Qualifiers
1..34593
/organism="Mus musculus"
/db_xref="taxon:10090"
join(<8470..8690,9077..9127,9266..9336,9948..10012,
10957..11014,12096..12164,12933..13064,14902..14960,
16609..16667,16746..16818,17930..17482,18671..18766,
21539..21677,21757..21842,23553..23671,24619..24715,
27291..27429,27763..27801,27930..28039,28553..28652,
28733..28853,29101..29179,29710..29854,30406..30639)
/gene="Elac2"
/note="alternatively spliced 24 exon form"
/product="putative prostate cancer susceptibility protein"
join(<8470..8690,9077..9127,9266..9336,9948..10012,
10957..11014,12096..12164,12933..13064,14902..14960,
16609..16667,16746..16818,17930..17482,18671..18766,
21539..21677,21757..21842,23553..23671,24619..24715,
27291..27429,27763..27801,27930..28039,28553..28652,
28733..28853,29101..29179,29710..29854,30406..30639)
/gene="Elac2"
/note="Elac2"
/note="alternatively spliced 25 exon form"
/product="putative prostate cancer susceptibility protein"
join(<8470..>31277
/gene="Elac2"

FEATURES
SOURCE

mRNA

mRNA

gene
CDS

join(<8470..8690,9077..9127,9266..9336,9948..10012,
10957..11014,12096..12164,12933..13064,14902..14960,
16609..16667,16746..16818,17930..17482,18671..18766,
21539..21677,21757..21842,23553..23671,24619..24715,
27291..27429,27763..27801,27930..28039,28553..28652,
28733..28853,29101..29179,29710..29854,30406..30639)
/gene="Elac2"
/note="24 exon form"
/product="putative prostate cancer susceptibility protein"

```
/protein_id="AAK29421.1"
/db_xref="GI:13540343"
/translation="MMALRSILRPLGLRTMSGSSARRPPPKDPLRLHLTRREKGRCP
GGPNTVYLQVAVAGDAGALYFSEYRYRYFNGCEGVQRLMOEKLKLVARDNLFTL
TMHWSNWGLGMITLTKETGLPKCVLSGPOLKEYLLEAIKIFSGPLGIELAVPH
SAPEYKDEMTYQVPIHSEKRCGKQSPSPRTSNRLSPKSSDSSGSAENGQCPPE
DSSAGANRRKMGDRPSLVYAFVCKLHLRKGNFLVKAKELGIPVGAALAPITIAAVK
GKSTVEGREIAAEELCPDPDGLVPIVECEDEGLIPICENDPFKRRQOARDAVA
LVYHIAPESVLIDSRQOMMERFPDTHLITENCPVYHNRSHKIQIOTLSLTHPDI
FQULTSFYSKEGSTLSPVTRGECILKYOLPRKRWQDTLDCNTDEFIAEALP
SFOESVEEYRKVQENPAPAEKRSQPEIYFGLTGSATIPMKIRNVSSTLVNLSPE
LIDCGEGFGCLRHGQOIDLRYLCSLTAIVFVSHLADHTGLNLTLREHNLASLG
LETCDELEEFQTCILVRCKHAFGCALVHSGMKVYSGDIMPCCALYQMGKDALILHE
ATLEDGLEEAEVEKTHSTSOAINVGMNNAEFTIMNHRSORAKIPLFSPDNEKVG
IAPDHMKVCGDPEYPKLIPLKAFACDIEYMRREKRELIVRALILTQOADSP
EDREPOQRAHDEPHSPQSKESVANTIGARY"
CDS
join(8470..8690,9077..9127,9266..9336,9948..10012,
10957..11014,12096..12164,12933..13064,14902..14960,
16609..16667,16746..16818,17370..17482,18671..18766,
21539..21677,21757..21842,23553..23671,24619..24715,
27291..27429,27763..27801,27930..28039,28553..28652,
28733..28853,29101..29179,29710..29854,30406..30624,
31242..31277)
/gene="Eiac2"
/note="25 exon form"
/codon_start=1
/product="putative prostate cancer susceptibility protein"
/protein_id="AAK29420.1"
/db_xref="GI:13540342"
/translation="MMALRSILRPLGLRTMSGSSARRPPPKDPLRLHLTRREKGRCP
GGPNTVYLQVAVAGDAGALYFSEYRYRYFNGCEGVQRLMOEKLKLVARDNLFTL
TMHWSNWGLGMITLTKETGLPKCVLSGPOLKEYLLEAIKIFSGPLGIELAVPH
SAPEYKDEMTYQVPIHSEKRCGKQSPSPRTSNRLSPKSSDSSGSAENGQCPPE
DSSAGANRRKMGDRPSLVYAFVCKLHLRKGNFLVKAKELGIPVGAALAPITIAAVK
GKSTVEGREIAAEELCPDPDGLVPIVECEDEGLIPICENDPFKRRQOARDAVA
LVYHIAPESVLIDSRQOMMERFPDTHLITENCPVYHNRSHKIQIOTLSLTHPDI
FQULTSFYSKEGSTLSPVTRGECILKYOLPRKRWQDTLDCNTDEFIAEALP
SFOESVEEYRKVQENPAPAEKRSQPEIYFGLTGSATIPMKIRNVSSTLVNLSPE
LIDCGEGFGCLRHGQOIDLRYLCSLTAIVFVSHLADHTGLNLTLREHNLASLG
LETCDELEEFQTCILVRCKHAFGCALVHSGMKVYSGDIMPCCALYQMGKDALILHE
ATLEDGLEEAEVEKTHSTSOAINVGMNNAEFTIMNHRSORAKIPLFSPDNEKVG
IAPDHMKVCGDPEYPKLIPLKAFACDIEYMRREKRELIVRALILTQOADSP
EDREPOQRAHDEPHSPQSKESVANTIGARY"
BASE COUNT      8641 a      8030 c      8142 g      9780 t
ORIGIN
Query Match      3.5%; Score 35; DB 94; Length 34593;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      891 gggaagatgcacccctcgtacatgaagccac 925
|||||
Db      29100 GCGAAGATGCCACCTCTGATCATGAAGCCAC 29134
|||||
RESULT 11
AF165142      105490 bp      DNA      PRI      24-AUG-1999
LOCUS      Homo sapiens chromosome 8 clone BAC 392C11 map 8p11-p12, complete
DEFINITION
sequence.
ACCESSION      AF165142
VERSION      AF165142.1      GI:5499748
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 105490)
AUTHORS      Schudy,A., Blechschmidt,K., Schillhabel,M., Baumgart,C., Menzel,U.,
Weber,J., Schattervoy,R. and Rosenthal,A.
```

TITLE	Direct Submission
JOURNAL	Submitted (05-JUL-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES	Location/Qualifiers
source	1..105490
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="8"
	/map="8p11-p12"
	/clone="BAC 392C11"
	/complement(84..140)
	/note="GRAIL"
exon	/evidence=not_experimental
	complement(333..592)
repeat_region	/rpt_family="AluX"
	716..761
repeat_region	/rpt_type=tandem
	complement(1299..1513)
repeat_region	/rpt_family="MIR"
	3417..3519
exon	/note="MZF"
	/evidence=not_experimental
	complement(3617..4219)
repeat_region	/rpt_family="Charliela"
	complement(4209..4252)
repeat_region	/rpt_family="Charliela"
	4302..4405
exon	/note="GRAIL"
	/evidence=not_experimental
	complement(5085..5193)
exon	/note="MZF"
	/evidence=not_experimental
	complement(5315..5460)
repeat_region	/rpt_family="FLAM_C"
	complement(5347..5460)
repeat_region	/rpt_family="AluO"
	complement(5373..5537)
exon	/note="MZF"
	/evidence=not_experimental
	6488..6564
exon	/note="GRAIL"
	/evidence=not_experimental
	complement(6760..7029)
repeat_region	/rpt_family="AluY"
	7180..7227
repeat_region	/rpt_type=tandem
	complement(7232..7489)
repeat_region	/rpt_family="AluX"
	complement(7489..7563)
exon	/note="GRAIL"
	/evidence=not_experimental
	8092..8142
exon	/note="Initial_exon"
	complement(8101..8149)
exon	/note="GRAIL"
	/evidence=not_experimental
	8306..8360
exon	/note="GRAIL"
	/evidence=not_experimental
	9096..9528
repeat_region	/rpt_family="Tigerza"
	9191..9229
repeat_region	/rpt_type=tandem
	9223..9282
exon	/note="GRAIL"
	/evidence=not_experimental
	9223..9282
exon	/note="Terminal_exon"
	10179..10635
repeat_region	/rpt_family="MER31B"
	10824..10879
repeat_region	/rpt_type=tandem
	10882..11145

```

repeat_region /rpt_family="MER31-internal"
11138. .11339
/rpt_family="MER31-internal"
repeat_region /rpt_family="MER31-internal"
11351. .11994
/rpt_family="MER31-internal"
repeat_region 11563. .11606
/rpt_type=tandem
repeat_region 12029. .12223
/rpt_family="MER57-internal"
repeat_region 12122. .12742
/rpt_family="MER31-internal"
repeat_region 12457. .13337
/rpt_family="MER4-internal"
repeat_region 12637. .13415
/rpt_family="MER65-internal"
repeat_region 12749. .14506
/rpt_family="MER31-internal"
exon 13011. .13027
/note="XPOUND"
/evidence=not_experimental
exon 14127. .14173
/note="XPOUND"
/evidence=not_experimental
exon 14239. .14331
/note="GRAIL"
/evidence=not_experimental
exon 14439. .14470
/note="XPOUND"
/evidence=not_experimental
exon 14513. .14985
/rpt_family="MER31B"
repeat_region 14986. .15205
/rpt_family="MER31-internal"
repeat_region 15583. .15624
/rpt_type=tandem
repeat_region 15594. .16226
/rpt_family="AluYa8"
repeat_region 15778. .16081
/rpt_family="AluSc"
repeat_region 16414. .16766
/rpt_family="THE1B"
repeat_region 17761. .18189
/rpt_family="L1PA13"
repeat_region 18043. .18271
/rpt_family="L1"
repeat_region 18286. .18554
/rpt_family="L1PA10"
exon 18291. .18397
/note="MZF"
/evidence=not_experimental
exon 18291. .18397
/note="GRAIL"
/evidence=not_experimental
repeat_region 18661. .19071
/rpt_family="MT1F"
repeat_region 19781. .20183
/rpt_family="MER83"
exon 20013. .20114
/note="GRAIL"
/evidence=not_experimental
repeat_region 21158. .21279
/rpt_family="MIR"
repeat_region 21802. .21906
/rpt_family="MER5A"
repeat_region 22020. .22325
/rpt_family="AluDb"
repeat_region 22911. .23014
/rpt_family="MT1J"
repeat_region 23103. .23196
/rpt_family="L2a"
exon 23581. .23725
/note="GRAIL"
/evidence=not_experimental

```

```

exon 23583. .23663
/note="GRAIL"
/evidence=not_experimental
exon 24026. .24185
/note="GRAIL"
/evidence=not_experimental
exon 24114. .24185
/note="Terminal_exon"
exon 24114. .24185
/note="XPOUND"
/evidence=not_experimental
exon 24201. .24344
/note="GRAIL"
/evidence=not_experimental
repeat_region 24684. .24945
/rpt_family="L1MA6"
exon 25524. .25621
/note="XPOUND"
/evidence=not_experimental
repeat_region 25642. .25699
/rpt_family="L2"
repeat_region 25837. .25924
/rpt_type=tandem
repeat_region 26272. .27029
/rpt_family="L1PA7"
exon 26733. .26881
/note="GRAIL"
/evidence=not_experimental
exon 27656. .27783
/note="GRAIL"
/evidence=not_experimental
repeat_region 2765. .28027
/rpt_family="LTR37B"
repeat_region 28150. .28450
/rpt_family="MER33"
repeat_region 28452. .29399
/rpt_family="L1MA5"
repeat_region 28897. .29188
/rpt_family="AluSc"
repeat_region 29402. .29593
/rpt_family="Charles5"

```

```

Query Match 2.3%; Score 23; DB 89; Length 105490;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 289 aagaagtcagatccagaataca 311
Db 39697 AGAAGTCAGTACCGAATCA 39719

```

```

RESULT 12
AC019331 190705 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 8 clone RP11-104D16, WORKING DRAFT
DEFINITION AC019331
ACCESSION AC019331
VERSION AC019331.4 GI:8569783
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 190705)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 190705)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

COMMENT On Jun 16, 2000 this sequence version replaced gi:7717162.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WOGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0104D16
----- Summary Statistics -----
Sequencing vector: M13; 498
Sequencing vector: plasmid; 518
Chemistry: Dye-primer ET; 49% of reads
Chemistry: Dye-terminator Big Dye; 51% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 182645 bases at least Q40
Consensus quality: 185230 bases at least Q30
Consensus quality: 186612 bases at least Q20
Insert size: 199000; agarose-fp
Insert size: 189205; sum-of-contigs
Quality coverage: 4.52 in Q20 bases; agarose-fp
Quality coverage: 4.79 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 2064: contig of 2064 bp in length
* 2065 2164: gap of unknown length
* 2165 3805: contig of 1641 bp in length
* 3806 3905: gap of unknown length
* 3906 8649: contig of 4744 bp in length
* 8650 8749: gap of unknown length
* 8750 13321: contig of 4572 bp in length
* 13322 13421: gap of unknown length
* 13422 19509: contig of 6088 bp in length
* 19510 19609: gap of unknown length
* 19610 24715: contig of 5106 bp in length
* 24716 24815: gap of unknown length
* 24816 30643: contig of 5828 bp in length
* 30644 30743: gap of unknown length
* 30744 40713: contig of 9970 bp in length
* 40714 40813: gap of unknown length
* 40814 48495: contig of 7682 bp in length
* 48496 48595: gap of unknown length
* 48596 57063: contig of 8468 bp in length
* 57064 57163: gap of unknown length
* 57164 67250: contig of 10087 bp in length
* 67251 67350: gap of unknown length
* 67351 82559: contig of 15209 bp in length
* 82560 82659: gap of unknown length
* 82660 102558: contig of 19899 bp in length
* 102559 102658: gap of unknown length
* 102659 124552: contig of 21894 bp in length
* 124553 124652: gap of unknown length
* 124653 152228: contig of 27576 bp in length
* 152229 152329: gap of unknown length
* 152329 190705: contig of 38377 bp in length.
Location/Qualifiers
1. 190705
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="RP11-104D16"
1. 2064
/misc_feature
/feature="assembly_name:Contig7"
2165. 3805
/feature="assembly_name:Contig8"
clone_end:SP6
vector_side:left"

```

```

misc_feature 3906. 8649
/feature="assembly_name:Contig9"
misc_feature 8750. 13321
/feature="assembly_name:Contig10"
misc_feature 13422. 19509
/feature="assembly_name:Contig11"
misc_feature 19610. 24715
/feature="assembly_name:Contig12"
misc_feature 24816. 30643
/feature="assembly_name:Contig13"
misc_feature 30744. 40713
/feature="assembly_name:Contig14"
misc_feature 40814. 48495
/feature="assembly_name:Contig15"
misc_feature 48596. 57063
/feature="assembly_name:Contig16"
misc_feature 57164. 67250
/feature="assembly_name:Contig17"
misc_feature 67351. 82559
/feature="assembly_name:Contig18"
misc_feature 82660. 102558
/feature="assembly_name:Contig19"
misc_feature 102659. 124552
/feature="assembly_name:Contig20"
misc_feature 124653. 152228
/feature="assembly_name:Contig21"
misc_feature 152329. 190705
/feature="assembly_name:Contig22"
BASE COUNT 55947 a 36657 c 35874 g 60716 t 1511 others
ORIGIN

```

```

Query Match 2.3%; Score 23; DB 65; Length 190705;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 289 aagaagtcagtcaccagaatca 311
|||||
Db 161565 AAGNAGTCAGTACCAGAAATCA 161587

```

```

RESULT 13
AL139399/c PRI 21-DEC-2000
DEFINITION
Human DNA sequence from clone RP11-574A21 on chromosome
AL139399
ACCESSION
AL139399
VERSION
AL139399.9 GI:11989998
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 64151)
REFERENCE
1. Submitted (21-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Dec 22, 2000 this sequence version replaced gi:11878435.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:

```

Em.: EMBL: Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone configs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

Rp11-574A21 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone Rp11-574A21 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone Rp11-274M8 is at 64052 in this sequence. The true right end of clone Rp1-267M20 is at 100 in this sequence.

FEATURES

location/Qualifiers

1..64151

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="q21.33-22.3"

/clone_lib="RPC1-11.2"

BASE COUNT 20073 a 11574 c 12196 g 20308 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 2; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 212 ctgcagctcccaactccag 232

|||||

DB 24576 CTGCAGCTTCCCACTCCAG 24556

RESULT 14

AB046207

LOCUS AB046207 1176 bp mRNA VRT 11-APR-2001

DEFINITION Anguilla japonica mRNA for 28Kda-1d apolipoprotein, complete cds.

AB046207

AB046207.1 GI:13591607

KEYWORDS

SOURCE

Anguilla japonica liver cDNA to mRNA.
Anguilla japonica
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Anguilliformes;

REFERENCE 1 (sites)
Kondo, H., Kawazoe, I., Nakaya, M., Kikuchi, K., Aida, K. and Watabe, S.

The novel sequences of major plasma apolipoproteins in the eel
Anguilla japonica(1)

Biochim. Biophys. Acta 1531 (1-2), 132-142 (2001)

JOURNML 21173592

REFERENCE 2 (bases 1 to 1176)
Watabe, S., Kondo, H., Kawazoe, I., Nakaya, M., Kikuchi, K. and Aida, K.

Direct Submission
Submitted (16-JUL-2000) Shugo Watabe, The University of Tokyo,
Graduate School of Agricultural and Life Sciences; Yayoi 1-1-1,
Bunkyo, Tokyo 113-8657, Japan
(E-mail: watabe@mail.ecc.u-tokyo.ac.jp, Tel: 81-3-5841-7520,
Fax: 81-3-5841-8166)

Location/Qualifiers

1..1176

/organism="Anguilla japonica"

/db_xref="taxon:7937"

/tissue_type="liver"

18..797

/codon_start=1
/product="28kda-1d apolipoprotein"
/protein_id="BAB40964.1"
/db_xref="GI:13591608"

CDS

/translation="MKFIVIALVSTFGCAQANLRSNEPTPOLELVKDAFWDYEVKAS
QTAAQALKTGESELAQQVNAKIKESVEVVOQRYRIVOEQIVISDELHKKLSHAQ
LSEISQDPIENEPVLAQLEPLAKLANIQOQKRPVDEQVPTESIDIALRLRRMH
RLRMVSVQLLSLEHLODOLGSGTEELKRRKESVQVEYREFPLPLENIGNELINK
LFMEFRVYPPVDCIRKREKLDPRYIGLGEQLTALWESSANSA"

BASE COUNT 333 a 262 c 286 g 295 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 7.9; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 234 agagcgtgcagagtcacag 253

|||||

DB 601 ACAGCGTCGACGATGACAG 620

RESULT 15

AFU23796

LOCUS AFU23796 1573 bp mRNA PLN 21-JUL-1995

DEFINITION Arabidopsis thaliana IAA-amino acid hydrolase homolog ILL2 (ILL2)

ACCSSION mRNA, complete cds.

VERSION U23796

KEYWORDS U23796.1 GI:902790

SOURCE

thale cress.
Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1573)
Bartel, B. and Fink, G.R.

IRL1, an amidohydrolase that releases active indole-3-acetic acid
from conjugates

Science 268 (5218), 1745-1748 (1995)

JOURNML 95312867

MEDLINE 2 (bases 1 to 1573)
Bartel, B. and Fink, G.R.

REFERENCE Direct Submission
Submitted (27-MAR-1995) Bonnie Bartel, Whitehead Institute, 9
Cambridge Center, Cambridge, MA 02142, USA
On Jul 21, 1995 this sequence version replaced gi:887788.

JOURNML Location/Qualifiers

COMMENT

FEATURES

source

CDS

1..1573
/organism="Arabidopsis thaliana"
/strain="landsberg erecta"
/db_xref="taxon:3702"
/chromosome="5"
1..1320

/gene="ILL2"
/note="IRL1-like gene; IAA-amino acid hydrolase homolog"

/codon_start=1
/product="ILL2"
/protein_id="AAC49016.1"

/db_xref="GI:902791"

/translation="MALNKLILSLFFOLLFLSVSSSPWIAEDTSOTLOTFLFAKS
PEVFDWMVKIRKTHENDELIGEELSLKRSLSLELIGTVRPVAVITGVYIGTG
EPPFVALRADMDALPIQGVWEWEKSKIPGMNACGHDGHTMLLGAKIIHERRHL
QGTVALIFQPAEELSGAKMKREGALKNVAFIGHLSAIPGKAASRGSTLAGA
GVEFAVITGSGIGHAIPHTIDPVVAASVLSIQQLVSRTPDLSKVVYVSVNGS
NAFWVIDPSITIGTILAFGTGFTQLOQVKEVITRQAAVHRCNANVNLTPNGREPMPP
TVNKKDLVKQFKRVYRDLGOEAFVEAPVNGSEDFSYFAETIPGHFSILGMODETNG
YASHSPLRYRINEDVLPYGAIIHSMNVQYLKKAASGVSGFREFL"

75..1320

/gene="ILL2"

BASE COUNT 452 a 332 c 363 g 426 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 7.9; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 565 ctttggaagccgcttcac 584
 ||||||||||||||||
 Db 609 CTTTGGAAAGCCCTTCAC 628

RESULT 16
 AF047031
 LOCUS Arabidopsis thaliana IAA-amino acid hydrolase (ILL1 and ILL2)
 DEFINITION genes, complete cds.
 ACCESSION AF047031
 VERSION AF047031.1 GI:2921828
 KEYWORDS thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 4918)
 Bartel,B. and Fink,G.R.
 IRL1, an amidohydrolase that releases active indole-3-acetic acid from conjugates
 Science 268 (5218), 1745-1748 (1995)
 2 (bases 1 to 4918)
 Davies,R.T., Goetz,D.H., Lasswell,J., Anderson,M.N. and Bartel,B.
 IAR3 encodes an auxin conjugate hydrolase from Arabidopsis
 Plant Cell 11 (3), 365-376 (1999)
 3 (bases 1 to 4918)
 Bartel,B.
 Direct Submission
 Submitted (08-FEB-1998) Biochemistry and Cell Biology, Rice University, 6100 S. Main St., Houston, TX 77005, USA
 Location/Qualifiers
 1. 4918
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /map="between nga129 and LFY3"
 join(<143. 496,605. 730,805. 1119,1201. 1323,1408. >1806)
 /gene="ILL1"
 /product="IAA-amino acid hydrolase"
 <143. .>1806
 /gene="ILL1"
 /note="ILR1-like gene 1"
 join(143. 496,605. 730,805. 1119,1201. 1323,1408. 1806)
 /gene="ILL1"
 /function="auxin conjugate hydrolase"
 /codon_start=1
 /product="IAA-amino acid hydrolase"
 /protein_id="AAC04865.1"
 /db_xref="GI:2921829"
 /translation="MALNNFLRQLLLLRSSSESPWIVAGVSRIPINFLPLASPEVFDSEWIRIRKIHENPELGEFEFTSKTRSELIDIGVYRPVAITG1IGTIGEPTFVALRADMALPIQEAWEHRSKNPKMAGCGHDGVAMILGAATILQGRHLQGVVLFQPADEGLSGAKMREGALKNEATLIGILSRTPFGKASLAGSPMAGAGAEAVITGKGAAIPIQHTIDPVVAASIVLSIQHVSRETQPSDKVTVKRVNGN AFNVIPDSITIGTILRAPFTGQLOQRKEIKQAAYHRCNASVNLAPNGMPPT VNNMDLYKFKKYVRDLGOEAFVEAVPEKGSDESYFAETIPGHFSLIGMODETNGY ASHSPHYRINEDVLPYGAAIHATMAVOYLKOKASKSVSGFHEEL."
 join(<2596. 2952,3121. 3246,3323. 3637,3718. 3840, 3931. >4329)
 /gene="ILL2"
 /product="IAA-amino acid hydrolase"
 <2596. >4329
 /gene="ILL2"
 /note="ILR1-like gene 2"
 join(2596. 2952,3121. 3246,3323. 3637,3718. 3840, 3931. >4329)
 /gene="ILL2"
 /product="KIAA0616 protein"

variation
 /gene="ILL2"
 /note="6 base insertion in ecotype Landsberg erecta"
 /replace="tgaatt"

BASE COUNT 1422 a 991 c 997 g 1508 t
 ORIGIN

Query Match 2.0%; Score 20; DB 12; Length 4918;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 565 ctttggaagccgcttcac 584
 ||||||||||||||||
 Db 3448 CTTTGGAAAGCCCTTCAC 3467

RESULT 17
 AB014516
 LOCUS Homo sapiens mRNA for KIAA0616 protein, partial cds.
 DEFINITION AB014516
 ACCESSION AB014516.1 GI:3327045
 VERSION
 KEYWORDS
 SOURCE Homo sapiens adult male brain cDNA to mRNA, clone_11b:pbluescriptII SK plus Clone:HG03623.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 6718)
 Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
 Direct Submission
 Submitted (26-MAY-1998) to the DDBJ/EMBL/GenBank databases. Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)
 2 (sites)
 REFERENCE
 AUTHORS Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
 Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro
 DNA Res. 5 (3), 169-176 (1998)
 98403880
 Location/Qualifiers
 1. 6718
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HG03623"
 /clone_lib="pbluescriptII SK plus"
 /dev_stage="adult"
 /sex="male"
 /tissue_type="brain"
 1. 1906
 /gene="KIAA0616"
 <1. 1906
 /gene="KIAA0616"
 /codon_start=2
 /product="KIAA0616 protein"

```

/protein_id="BAA31591.1"
/db_xref="GI:3327046"
/translation="HNKQAERTAAFEVYMKDLSITRARLDLQKSOYLQGPSKRG
IGGSLEPNVNIQSGTMDLPFGPSGLGEHLAAAPSLTPEFSSGLDTSRTYHHHLIV
RYRRRGKSGSPHRRPLVLDKHQRQADSCPYGTMYLSIPADISMRNTSDSLHGST
TPTPEPSFSSGSDVHQRKRVLLITVPGMEETSEADKMLSKQAMDTKKGSPKCE
PGIINFPSPADODENTYALIPATYHNTGSLPDLTINIHFPSPLPRLDEEPTFALS
STGNLAANULTHLIGIGAGOGMSTPGSSOHRPAGVSPSLSTEARRQOAPPLISL
ITQVAMDALSLTEQLPYAFFTQASQDPPROPAGPPPPASQPPPPPPPPQAVR
PRGGLLEPSALSTRGPQPPPLAVTYPSSLPOSPPENPGPSMGITIASPAAQYRT
AGSPANQSPSTSPVSNQGFSPGSSPQHSTYTLGSEFEDAYEQQAAQAALSHOLEQ
NMMEVAISSSSLYSPGSTLNSQAAMMGITLSPQOGLVASHGSIPIILTV
GESPSLSEKLETLTSLAGVDVDSFDSDSOFLDELKIDPLTLDGLHMLNDPMDVADP
TEDIETRMRL"
BASE COUNT      1208 a      2237 c      1930 g      1343 t
ORIGIN
Query Match      2.0%; Score 20; DB 85; Length 6718;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Oy      462 5ggtctctggacacctgacctgct 481
         |||||||
Db      3731 GGGTCTCTGGGACACCTTGCT 3750

RESULT 18
AC003107 LOCUS      AC003107      46275 bp      DNA      PRI      18-NOV-1997
DEFINITION      Human DNA from chromosome 19-specific cosmid R30064 containing the
ACCESSION      AC003107
VERSION      AC003107.1
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 46275)
AUTHORS      Lamerdin,J.E., McCreedy,P.M., Adamson,A.W., Burkhardt,Schultz,K.,
Gordon,L., Kyle,A., Ramirez,M., Stillwagen,S., Garges,J.,
Dangnan,L., Christensen,M., Bruce,R., Quan,G., Montgomery,M.,
Ow,D., Kobayashi,A., Olsen,A.O. and Carrano,A.V.
TITLE      Sequence analysis of an -1 Mb region containing the MEF2B gene in
19p12
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 46275)
TITLE      Lamerdin,J.E.
JOURNAL      Direct Submission
COMMENT      Submitted (18-NOV-1997) Human Genome Center, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
Cosmid R30064 overlaps cosmid F19807 to the left and cosmid R32469
to the right.

FEATURES
Source
Map and sequence oriented from telomere to centromere.
Location/Qualifiers
1..46275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R30064"
/chromosome="19"
/map="19p12 between UBA52 and D19S451"
/cell_line="5H12-B"
/clone_id="HL19NC03 R chromosome 19 cosmid library"
/note="HL19NC03 cosmid library constructed at LNL from
flow-sorted chromosomes from hybrid 5H12-B, which carries
chromosome 19 as its only human chromosome."
817..1037
/note="predicted exon, program: gail2exons_human_1.3,
frame: 1, quality: excellent, score: 93.000"
complement(3847..3899)
/rpt_family="MER7"
repeat_region

```

```

repeat_region complement(3910. .4198)
                /rpt_family="Alu"
repeat_region complement(4214. .4315)
                /rpt_family="MER7"
misc_feature 6213. .6337
                /note="predicted exon, program: grai12exons_human_1.3,
                frame: 0, quality: excellent, score: 93.000"
repeat_region 6488. .6785
                /rpt_family="Alu"
repeat_region complement(6934. .7242)
                /rpt_family="Alu"
repeat_region complement(7269. .7545)
                /rpt_family="Alu"
repeat_region complement(7989. .8304)
                /rpt_family="Alu"
misc_feature 9294. .9409
                /note="predicted exon, program: grai12exons_human_1.3,
                frame: 2, quality: good, score: 65.000"
misc_feature 9543. .9602
                /note="DDS similarity to AA378771 EST91505 Synovial
                sarcoma Homo sapiens cDNA 5' end (1. .60); 100% identity."
misc_feature 9676. .10030
                /note="DDS similarity to AA209234 zgb5f01.r1 Stratagene
                hNT neuron (#937233) Homo sapiens cDNA clone 648409 5'
                (1. .351); Score: 660 Identity: 346/351 (98%)."
misc_feature complement(10340. .11045)
                /note="DDS similarity to AA211938 zgb5f01.s1 Stratagene
                hNT neuron (#937233) Homo sapiens cDNA clone 648409 3'
                Score: 951 Identity: 497/501 (99%)."
misc_feature 11916. .12187
                /note="DDS similarity to T82171 yd95906.r1 Homo sapiens
                cDNA clone 116026 5' similar to contains TAR1 repetitive
                element.Score: 478 Identity: 262/274 (95%)."

Other overlapping matches:
(11993. .12136) predicted exon, program:
grai12exons_human_1.3, frame: 2, quality: excellent,
score: 81.000
(12176. .11917) DDS similarity to T95404 ye43901.r1 Homo
sapiens cDNA clone 120528 5'. Score: 471 Identity:
258/263 (98%)."(12713. .11917) DDS similarity to T84000
yd6610.r1 Homo sapiens cDNA clone 113203 5'. Score: 400
Identity: 257/275 (93%)."(11965. .12181) DDS similarity to
T95320 ye43901.s1 Homo sapiens cDNA clone 120528 3'
similar to gb:M64241 QM PROTEIN (HUMAN).Score: 334
Identity: 203/218 (93%)."
misc_feature 12251. .12355
                /note="predicted exon, program: grai12exons_human_1.3,
                frame: 1, quality: good, score: 52.000-DDS similarity to
                AA378771 EST91505 Synovial sarcoma Homo sapiens cDNA 5'
                end (61. .165). 96% identity."
misc_feature 13676. .13705
                /note="predicted exon, program: grai12exons_human_1.3,
                frame: 1, quality: good, score: 54.000"
misc_feature 14355. .14457
                /note="predicted exon, program: grai12exons_human_1.3,
                frame: 0, quality: excellent, score: 77.000"
misc_feature 15674. .15795
                /note="predicted exon, program: grai12exons_human_1.3,
                frame: 0, quality: good, score: 65.000"
misc_feature 15709. .15764
                /note="DDS similarity to AA378771 EST91505 Synovial
                sarcoma Homo sapiens cDNA 5' end (166. .221); 99%
                identity."
misc_feature 16450. .16630
                /note="predicted exon, program: grai12exons_human_1.3,
                frame: 0, quality: excellent, score: 100.000"
misc_feature 17229. .17316
                /note="DDS similarity to T95918 ye42404.r1 Homo sapiens
                cDNA clone 120366 5'. Score: 163 Identity: 86/87 (98%)."
misc_feature 18023. .18187

```

```

/misc_feature
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 100.000"
18533..18757
/note="DSS similarity to AA252605 zsl4p02.sl NCI_CGAP_GCB1
Homo sapiens cDNA clone 685131 3' end. Score: 424
identity: 225/225 (100%)."
misc_feature
/note="DSS similarity to overlapping ESTs:
(20682..22267) AA350958 EST58512 Infant brain Homo sapiens
cDNA 5' end. Score: 776 identity: 391/394 (99%)
(20863..21259) R87396 ym88h07.r1 Homo sapiens cDNA clone
166045 5' end. Score: 748 identity: 391/401 (97%)
(20874..21249) AA349930 EST57030 Infant brain Homo sapiens
cDNA 5' end. Score: 748 identity: 375/376 (99%)
(21125..21574) R19558 y926h07.r1 Homo sapiens cDNA clone
33726 5' end. Score: 811 identity: 440/457 (96%)
(21589..22030) H30788 y079106.r1 Homo sapiens cDNA clone
184163 5' similar to gb:M33326 NONSPECIFIC CROSS-REACTING
ANTIGEN NCA-95 (HUMAN);contains Alu repetitive element.
Score: 817 identity: 433/443 (97%)
(22007..22267) AA32326 EST4948 Cerebellum II Homo
sapiens cDNA 5' end. Score: 479 identity: 252/262 (96%)
"
repeat_region
complement(21512..21626)
/rpt_family="Alu"
misc_feature
/note="DSS similarity to R88263 ym90h09.sl Homo sapiens
cDNA clone 166241 3' end. Score: 837 identity: 455/478
(95%).-Other overlapping matches:
(23013..22497) H05926 y171f07.sl Homo sapiens cDNA clone
43533 3' end. Score: 872 identity: 497/527 (94%)
(22971..22534) AA46916 zw85a01.sl Scores fetal heart
NB2HR8 9w Homo sapiens cDNA clone 783720 3' end. Score: 857
identity: 436/437 (99%)
(22971..22561) W95490 ze02g07.sl Scores fetal heart NBH19W
Homo sapiens cDNA clone 357852 3' end. Score: 774 identity: 399/411 (97%)
(22566..22949) W95528 ze02g07.r1 Scores fetal heart
NBH19W Homo sapiens cDNA clone 357852 5' end. Score: 674
identity: 366/382 (95%)
(22996..22634) T33668 EST58674 Homo sapiens cDNA 3' end
similar to None.. Score: 710 identity: 359/363 (98%)."
complement(23576..23770)
/note="DSS similarity to N94385 zb76d06.sl Scores
senescent fibroblasts NbHSF Homo sapiens cDNA clone 309515
3' similar to gb:219585 THROMBOSPONDIN 4 PRECURSOR (HUMAN);
(194..1): 92% identity.-Other overlapping matches:
(23724..23770) predicted exon, program:
grail2exons_human_1.3, frame: 1, quality: good, score:
74.000-(23770..23664) DSS similarity to M36954 mb82e10.r1
Scores mouse p3NMF19.5 Mus musculus cDNA clone 335946 5'
similar to
gb:219585 THROMBOSPONDIN 4 PRECURSOR (HUMAN); gb:M86620
Mouse thrombospondin 3 (MOUSE) (307..423): 63% identity."
complement(23683..32077)
/gene="COMP"
/note="CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR
(COMP)"
CDS
complement(join(23724..23770,23683..24002,25000..25172,
25705..25901,26307..26355,26482..26660,26774..26955,
27048..27100,27341..27459,28299..28458,29070..29127,
29218..29322,29400..29558,29647..29721,29968..30105,
30750..30922,31370..31421,31658..31743,31999..32077))
/gene="COMP"
/note="CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR
(COMP)"
/codon_start=1
/product="COMP_HUMAN"
/protein_id="AAB86501.1"
/db_xref="GI:2623750"
/translation="MVPDYACVLLLTALAGAGGQSGPLGSDLPQMLREIQETNAA
LDVRELLRQVREITFLKNTWMECDACGMOQSVRTGLPSVRLHLHCARGCFPGVAC

```

```

FEATURES
    source
        1..79590
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /chromosome="5"

Query Match
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2.0%; Score 20; DB 85; Length 46275;

10TESGARGCPAGFTNGSHCTDVNECNAAHPCPRVRCINTSPGFRCEACPGYSG
PHQVGLFAKANKOVCDINECEGNCVCNPNSCAITRSCFGCPGPDVDAQ
SGCORAORFRCPPSCBCEHADVCLEEDSGRSRCVAGMAGNGITLCGRDLDLGF
DKRIKCEPERCKKNCVCVYVNSGQEDVPDVGICDACPADDDGVPNEKNCPLVNP
DORNTDEKMGACDCNCRSOKKNDKQDTQDGRFGACDSDDDQDGDGHQDSRDCPT
VNSAQEDSDHSGDGDACDDDDDDNGVPSRNCRLVNPNGQEDADRDSVGVGDDF
DADKVDKIDVCPENAEVTLTDFRAVFTQVLPDPEGQAQIDPMWVYLNQGREIYQTMNS
DGLAVGYAFNAGVPEGEFHNVTYTDQDYGAGFTEGQDSSFFVYVMKQMDQYQWA
NFRVAVEGIQLKAVKSSGTGCBQLRNALHMTGTDESGVRLMDPRVWKKDKKSY

```

```

RESULT 19
AB007727
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui PI
clone:MXC9.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
AUTHORS
Nakamura,Y., Sato,S., Kaneko,T., Kotani,H., Asamizu,E., Miyajima,N.
and Tabata,S.
TITLE
Structural analysis of Arabidopsis thaliana chromosome 5. III.
Sequence features of the regions of 1,191,918 bp covered by
seven physical clones
JOURNAL
DNA Res. 4 (6), 401-414 (1997)
MEDLINE
98162728
REFERENCE
2 (bases 1 to 79590)
Nakamura,Y.
Direct Submission
JOURNAL
Submitted (06-OCT-1997) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:yinakam@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
COMMENT
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd-graph.cgi?c=MXC9
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremml.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.

```

CDS
/clone="MXC9"
/clone_lib="Mitsui P1"
complement(3423..3875)
/note="contains similarity to SI self-incompatibility protein
gene_id:MXC9.1"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB10024.1"
/db_xref="GI:9759373"
/translation="MKNLKQVHVYVYLIIQIAFSQVKGPSDLKMSLKSVYITNRLG
LGGDLTLNCHKSAADDGLKTLAPNGSMSEKRFSTVIGTLLFYCHFTWPGQSKRFDI
YDDRDGVSRHSICINCIMDISIGPCMSFSDDAFNICYDMNGMLRT"
complement(5454..5903)
/note="gene_id:MXC9.2
unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB10025.1"
/db_xref="GI:9759374"
/translation="MLKTOVHVYVYLIIQIAFSQVKGPSDLKMSLKSVYITNRLG
DGSSTLCHKSAADDGLKTLAPNGSMSEKRFSTVIGTLLFYCHFTWPGQSKMNIY
DDRDGVSMGIPCLICIMDICKYCPCRSEIDDANITCYDMNGNRRS"
complement(join(6686..6926,7008..7210,7299..7592,
7870..7993,8085..9427))
/note="emb|CAB87679.1
gene_id:MXC9.3
strong similarity to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB10026.1"
/db_xref="GI:9759375"
/translation="MAEOKSSNGGGGGDVVINVVEBASRRKREMASPESBKGVPFS
KSPSEISKIVGSPMKPPRAPNQNNGVLFQKRSFARSVYSKPSRFVDPSPVDNISL
EEVEEOGACGSPSPRASPNKNSNSGSPAPYPSKYVEDEDEBETIKKKYLIREM
RSKISTLALISAFPVITLSALVASLTINVLKHHFTWGLEWAKCVLWAFVSGLVLT
NWMRLIVLELTNLFLLRKVLYFVHGLKSYQVFWICLLVAMILLFNHDKVSPA
ATVKLCITRRLISLTGAFPMVKTLLKILAAFNVNNEFDRIODSVFHOYVQTL
SGPLMEAEERVGSPSTGSHLSFATVVKKYGVEKKVIMDKGVHKMRKREKVSAMTRV
LMEAVRTSGLSITDLETAYEGKEBOADREITSEMAALAAHYVFNVAPOPFNI
EEEDLRPMKIEVDLVPFLPDGAETGRTRKAPTEGVVYITRRLAISINDTKT
AVKQNLKVTALIMVYVYVIMLLEVAITKYLEFSTOLVALAETIGSTCKNLDESI
VVEVFMHPIYDGRCDVGVDAVLEEMNLITVYFLKLNKAVYVNAVALAKRPIISNF
RSPNGETVEVSISFSTPVSKIAHLKERIAEYLAEONPOHMAVHSHVVKETEENMKL
MALYSDHTITFOENRERNLRTELSLAIKRMLDHDIDYTLIPODINLTKN"
complement(13439..17695)
/note="gene_id:MXC9.4"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="retroelement pol polyprotein-like"
join(18207..18234,18300..19381)
/note="contains similarity to NPk1-related protein kinase
gene_id:MXC9.5"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB10027.1"
/db_xref="GI:9759376"
/translation="MIENDSPQHSLEFDTFRANSYRVLHKKFLCKGYGSDLI
RYTIDGSSLOAAVSTYAEDELETDALKREIQIILSELKGYENIVICGDDLEEDFNE
HGHWYKLLLEYANEGSLSSPMENYDPDKRLDPMLRDTMLNGLVSHGSHVACH
LKSNDLILFSRRKDSASCELKIFEDGNCQVGEVDPHMKSDYVGTPESEFGVAKAT
LDWISGLCVLKITGTGEOPERVSVDFVNFSLDGEADNIPFYVCDAREFTECFAR
EHEKRTASELLIHPFLCOKOPILYKIKLRIKQASKITMGIRERKLTITKMLSKRPOL
KKLSKLRKKMTHPKPLELSDSFVSTSSDCLL"
19892..22342
/note="gb|AAFI9552.1
gene_id:MXC9.6
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB10028.1"
/db_xref="GI:9759377"

CDS
/translation="MYTRLVSRSSRYATVKFTDSVASCSCRLFSASTDPEPESOP
EOAPPNPVTGDEKLRNLRYVLIQONRIETARGVLSLRSSTFSPASKEFASFSLS
SPSLKHDFSVLTLVILNESKMISEADLPALINREGIYSPSSDLTLLDLVTKQF
RVTINVFNLITESPFRSPKPMYGRALIOAAVRLSPGKGLFELFNMRKHDRIVSEIYN
VLIDGICKGRMNABEOLFDEMLARLLPSLIYTNLIDGCKCANPKSPKVRERK
ADHTEPSLITFNTLTKGLFKAGVEDANVLEKMDLDFVPDAFTFTSLDFGYSNEK
AEALGVYETAVDSGVKNMAYTCSILLNALCKEKE TEKAEELIGREMAKGLVPEVYI
NTMIDGVCYRKADLVGARMKIEAMKQGRPHLYANCLIRPCELGEMENAKENYKM
KLKGVSPVETTYNLIIGYGRKYEDKCFDLKEMEDJMPNVSYGTLLNCLCKGS
KLBAQIVYKRDMDRGVSPKYRIYNNMIDGCCSKRGKIEDARFRKEMLKGIENLYT
YNTLIDGSMTGKISEADLLEISRKALKADVETYSNLISGYCFAGVQVCIALYEB
MKRSGIKPTKTLHLISLCTREGELETERJFGEKSLPDLVYNGVILCAVAGDME
KAFNLQKQMIKESIGLCKTNTYNSLILGQKVKGLCEVSSLIDENNAIREAEADYNI
IVKGCVEAKVDMSAVVWYREMQEGRFLDVCIGHELVISGLKEEKRSEAEIYISEMG
RMLGVTVDEDLSTEXL"
join(22987..23072,23172..23288,23378..23453,23623..23752,
23874..23987,24073..24236)
/note="emb|CAB64729.1
gene_id:MXC9.7"
/codon_start=1
/evidence-not_experimental
/product="elongation factor 1B alpha-subunit"
/protein_id="BAB10029.1"
/db_xref="GI:9759378"
/translation="MAVYFSDLHTERGILKTLLEHLAKTYSIGDQLSYDVKYAAVL
ENPDGFPNASKWYDVASHLAKSEFGKADVRVGGVAPPSSEAHPTHEEPADGDG
DDDDIDLEADETEDEKKAEEERAAKQFTKTKSGKSVYLEVKKPMDEFTDKKLE
AVRVSQMPGLTWGASKLVPGVGIKKLILMWTYIDDLVSVNLIEDHTISPPNNEYIQ
SVDIYAFNKIT"
complement(24733..26592)
/note="gene_id:MXC9.8
pI||T02643
strong similarity to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB10030.1"
/db_xref="GI:9759379"
/translation="MSSASKKMSBDKRYKMYNDSQKTPSKASGMSGSSVYVGYNPLLG
TFQRTESLATSGLSSLHNNGRFRHIEDSDSGACDSSNNGSSGEEDKEKVPST
SLAKQETILPADNDKDKMKLKNRKHQROKREKADLHRCQYLKMSRLEVAOL
ILISGIPHERATYALMLNEKRTIESINMLPDDGCAKADKLDPTSGKLKIDTIOELG
RILELRTYKCTKQDVERAVYTAGDELEKAEALRQKQDQSTASRKREDISDSTVN
NSKVPSTYTSQNTYGOQLDPNSGMYPAGEEELADKRNGLYPGSSYISGESNQSVNL
RIYHKKLDQMLQONAPLEENKRIYQOQTPLSRSTEEHYVYALGDQFKRLQDDMR
PVMVPMQOQOQOQOQOQANTVLLPVSTWNAFTGAAGAAGSMPHANSSEAAQSGNYLP
SRSLPNDLINSNLWYQOLOYOQYGVNNGHMAAGSSAPAVAPASLIGLFSGSL
AGSSSSGLDWNTPDSGLHDYNNIDMSLDKLCACVRBQCYVAASPTSAASPYEAM
MNGRTMNTNGMGVANGVGOEALVGVGREWTSFEBKDLFSLSKOTVPPSL"
join(28348..28450,28599..28744,28826..28902,28993..29080,
29161..29241,29336..29392,29485..29566,29659..29708,
29854..29910,29996..30074,30190..30248,30406..30463,
30561..30658)
/note="contains similarity to transmembrane transport protein
gene_id:MXC9.9"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB10031.1"
/db_xref="GI:9759380"
/translation="MLKCEVWTIISANRRLSHLPPIACSRGIDQDEDEKESRLLPH
KNDNATNRSRSSVDSGLNDYOEEERYKTSFTVALCVGTAAFGIGLKGAVK
TASEFPAGYILEQSSVDNLEFVYVLFKFKPLMYONKVLGYAAGIVFETLLILG
LATLOKFEAVNLAAVYLSFRLFASEEDDTLDNFIYKTCORFIPVSYSDGNR
FPTKHGDIKATPELLIYAVYLESDIAPANDSIPAVGVDPDPITVITSNFIATIGLR
SLYTLISRGDELEIYLOPSIAVYVGLGVKAKILDFRGHISTEASLSGVALLSTGVL
LSLTNKSSDS"
complement(join(31799..31984,32321..32440))
/note="emb|CAA03929.1
gene_id:MXC9.10"
/codon_start=1
/evidence-not_experimental
/product="cystatin"
/protein_id="BAB10032.1"

ILFPTLLYNLVRFKLQSFRCYAVDYRNFLEFKISEXQFSGDESMQKDVPRLLKTL
GVGVITLLNEPEYETLPPSLYSAYEMENHVIPTRDYLPAPSIDITLAVNFTHKALL
GKTYVHCAGRGSRSTTVLYCLYIEKHSMTVAAREHVSIRPVLILHPSQKVSILDR
FIDMLHQRNRTTKMT"
complement(join(16726..16824,17130..17372,17821..18332,
18890..19082,19231..19344))
/note="gene_id:MIK19..7
pir||T13434
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09880.1"
/db_xref="GI:1017676"

/translation="MKNKCNLIDSKLEENHNLGSKHPCGCGRMIOATKPNVGLP
AGVKEPDPOELIEHLKAKGKEKKKSSSHPLIDEPIITIDEDGICYHNPQKL
GVTRDGLSKHFFKSPRAATGTGRRRKTIORDHDSSELTSSETMHTKGRPMIN
GQORCKCKTLVYTNPKRRREPRTKNWQHQLHLINEEREGELVSKITPOTPRO
CVSNTHSDHSGSKDVIGIGVDEISVAVALQSLGSDVSVRVNMPHPTKSFDEGTA
EASKRGNOHVSIGCEVVDGIIITSSMSHHNIHDHNOHNOIGRERFHMSSYPM
PTTSQHSIFPHVTSTMPQSGVWFGIJRPVYGLHSYVRRRFLMLCI"
join(22516..22617,22808..22854,22935..22974,23115..23324,
23408..23566,23656..23709,23800..23998,24082..24344,
24431..24631,24742..24933)
/note="gene_id:MIK19..8"
/codon_start=1
/evidence=not_experimental
/product="pyrophosphate-dependent
phosphofructo-1-kinase-like protein"
/protein_id="BAB09881.1"
/db_xref="GI:1017676"

/translation="MSSPRSNPKIYNGPGYILQDVPHLIDYLPDPTYPNPLQDNP
ASVAVKQFVHADSDVEKRVYVHKGDPGRVRRAGPQKRVFESEDEVACIVTGG
CGLMVIVREYVSSLYMGVGRKILGIDVLSGTYGFAKNTIPLNSKVNINHGG
TIGSGRCHDPTNKIYDSTQDRIQNYIIGSDGRASGVFEETRRRLKVAAYGI
PRTINDIPYIDKSGFDIAYEADQAINAAHVBNSNGIGFKLMKRTISGITAM
ATLASRDVCCCLIPESPTLYEGEGLEFETERRLKDGHMVTIVLGAQODIMCKSME
SPMDASQCKLKLKDVGLMSQIKDHFKNKVMMLKYIDPTVMIRAVPSNADSVVC
TLAASAVGAMAGYGTGSLVNGROTPIPYRTLETENNVVITDRMARLLSTNO
PSFLGPDHSEKKELPERPLLDGAVDIPYTKETVK"
complement(join(25739..25765,25861..25942,26048..26126,
26210..26279,26372..26494,26652..26704,26807..26866,
26966..27170,27390..27474,27561..27690,27813..27843))
/note="gene_id:MIK19..9
pir||T06010
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09882.1"
/db_xref="GI:1017676"

/translation="MNISENVFVHEDSTQGTGLRLSDIPMSKISSDEVEFLAP
ENNAFGROPDYTDNSENROKSEVHEVYATQHTNOLDYOKRSEYGRKDKVMNIME
CCELSKYVDESDDPLDDEPOLOHLOLSAATIKDYPNEDMLHTALIDKCVLILPO
FGCLPQMAVYGTFFVGCANFDSNVHKKFMENPDPPNNKTYTKAGITSSEGGLENL
MSMGHDVYLVAKENGSTLPSPLGFLIRKHSFYPLHKAAGATYTHLMNEDEKLNKLWL
VNNKYDLSKSVHNVNEKVPYMSLIRKYPEENLRM"
join(29030..29383,29492..29617,29692..30006,30088..30210,
30295..30693)
/note="gene_id:MIK19..10
sp|P54969"
/codon_start=1
/evidence=not_experimental
/product="1Aa-amino acid hydrolase homolog 1 precursor"
/protein_id="BAB09883.1"
/db_xref="GI:10176769"

/translation="WALNFEITFQULLLLLRSSSPWIVADVSRIPINIELAKSP
EVEDSMVRIIRKHENPELGYEEFETSKIRSELDLIGKYRFPALITGIIYITGGE
PPFVILRADMDALPIQEAWEHESKNPCKMACHGCHDGVAVALLGAALIQOHRQL
GTVVLIQPAEGLSGAKMREKALKNVEALFGJHLSTPFGKAASLAGSEFMAGAG
AFAVITGKGHAIPQHTIDPVVAASIVLSLQHVSEETDPSDSKVYTVKAVNGN

Query Match 2.0%; Score 20; DB 12; Length 84129;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 565 cttggaaagccgcttcac 584
Db 32335 CTTGGGAAAGCCGCTTCAC 32354

RESULT 21
ATF14F18
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 103638)
Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A.,
Duesterhoeft,A., Bancroft,I., Mewes,H.W., Rudd,S., Lemcke,K. and
Mayer,K.F.X.
Unpublished
2 (bases 1 to 103638)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (13-APR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bsrc.ac.uk

COMMENT
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES
source
Location/Qualifiers
1..103638
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="5"

misc_feature

1..8764
/note="Overlap with BAC T22P22. Please refer to this BAC
for annotation and analysis"

complement(8151..8375)
/gene="F14F18_10"

exon
complement(1
/number=1
complement(join(8151..8375,8495..8551,8650..8678,
8890..9346))
/gene="F14F18_10"

gene
8151..9346
/gene="F14F18_10"

gene
complement(join(8151..8375,8495..8551,8650..8678,
8890..9346))
/gene="F14F18_10"

CDS
/note="similarity to predicted proteins in *Synechococcus*,
Cyanophora and other organisms"
/codon_start=1
/product="putative protein"
/protein_id="CAB87657.1"
/db_xref="GI:7573351"

/translation="MAATVIGHFSPITSLKPPSRHRRRLHHNLSPKTPLOSLIKFG
AENSRRPPRRPLPFDCCVPPPOQPINXQSTSFSPMSGDLIERYSTPLFGAS
FAFVGLFVSWFGSIGPEYKRLILASSSIFVYTVLAVRMVIGMAYVDEGTGYD
GQWVWTEVVALDRLLSFSVKPVLAARKNTLVILVLSLIVNLGDSPIATSYRTY
RDPDRSSLPITGAYNDETARTFEPAFCGEPSSDL"
complement(8376..8494)
/number=1
complement(8495..8551)
/gene="F14F18_10"

exon
complement(8552..8649)
intron

```
/number=2
complement(8650..8678)
/feature="F14F18_10"
exon
/number=3
complement(8679..8889)
/number=3
complement(8890..9346)
/feature="F14F18_10"
exon
/number=4
complement(join(11457..11637,11711..11760,11866..12033,
12133..12211,12407..12507,12606..12674,12823..12863,
13089..13356,13438..14056,14134..14237,14416..14856,
14953..15048,15122..15565))
/feature="F14F18_20"
exon
complement(11457..11637)
/feature="F14F18_20"
/number=1
11457..15565
/feature="F14F18_20"
complement(join(11457..11637,11711..11760,11866..12033,
12133..12211,12407..12507,12606..12674,12823..12863,
13089..13356,13438..14056,14134..14237,14416..14856,
14953..15048,15122..15565))
/feature="F14F18_20"
/feature="The mRNA covering this gene is only partial. This
has allowed the definition of the carboxy terminus of the
protein. The amino terminus remains predicted but
prediction is supported by different gene models."
/codon_start=1
/product="MAP3k delta-1 protein kinase"
/protein_id="CAB87658.1"
/db_xref="GI:7573352"
/translation="MSKKHLLRKHLTGSSGCGGFADHHRLDSTPMIDPILLS
ISPSDSTSVSSSGGNGASTMPRLDTEPVGRDITAVDGVDFNLMEBEYQVQLAMA
ISVSDPRENADTQDLDAKRLSLGVSAPVVDADSAVDLRLRYSGSISGLKEL
GSMILPAYSSSLRVFLPNMGKHLVINYDQKRVDFGVDYGTNSLSQKMLVLDLQ
AISLDNVYEVILVNLIDPELOELERFVFLASCPGPAPQGVSSDLOKLANIVY
EOMGGVENADALRRMRLRSYELRNSLTNTLIPLRGVAVGLARHALLFKVLAIVY
LPCMLYKGSYYTGTDGAVNLKLDKSPNSFTGCEYIIDLMGAPCALIPSPVPSF
LVSCTDTRVPEPNDLSLOHSSVLEKELETPAFSVSKPADSRSGVAVFTGNQEN
SRCVAVEKHQTERFEHDFGKLMHSQIISENMPFSGKPTCAQKRVKNSVISA
KNPEFAOKLHAVLESGASPPDLFMDINDPNHLRGKNLQELROESSNSMVGIPCY
EVARLDRESERNPAESYQOQSEVDLSMKRNPDLDTGKASSSEMEGTADGESAV
CDSHOGIINPLIGEAKEIMMEDLOIGERIGISGYCYRAAMNCTEYAVYKKFLDOD
ESGDALTOFKSEIETMLRLRHPNVILFKGAVTRPPNFTLTFELPRGSLYRLHRRNH
QUDKERRMMLDVAAGMNYLHTSHPTVVRHDKSPNLLVDKRMVAVCDPFGLSRKH
HYLLSSKSTAGTPPEWMAPEVLRNPEANEKCDVYSFGVILMEIATSHVPMKGLNPMQVH
GAVGFQNRLEIIPDDIDLTVAQIIRRCMQT"
complement(11638..11710)
/number=1
complement(11711..11760)
/feature="F14F18_20"
/number=2
complement(11761..11865)
/number=2
complement(11866..12033)
/feature="F14F18_20"
/number=3
complement(12034..12132)
/number=3
complement(12133..12211)
/feature="F14F18_20"
/number=4
complement(12212..12406)
/number=4
complement(12407..12507)
/feature="F14F18_20"
/number=5
complement(12508..12605)
/number=5
complement(12606..12674)
/feature="F14F18_20"
/number=6
complement(12675..12822)
/number=6
complement(12823..12863)
/feature="F14F18_20"
/number=7
complement(12864..13088)
/number=7
complement(13089..13356)
/feature="F14F18_20"
/number=8
complement(13357..13437)
/number=8
complement(13438..14056)
/feature="F14F18_20"
/number=9
complement(14057..14133)
/number=9
complement(14134..14237)
/feature="F14F18_20"
/number=10
complement(14238..14415)
/number=10
complement(14416..14856)
/feature="F14F18_20"
/number=11
complement(14857..14952)
/number=11
complement(14953..15048)
/feature="F14F18_20"
/number=12
complement(15049..15121)
/number=12
complement(15122..15565)
/feature="F14F18_20"
/number=13
15565..17936
/feature="F14F18_30"
complement(join(16565..16786,17110..17447,17579..17936))
/feature="F14F18_30"
complement(join(16565..16786,17110..17447,17579..17936))
/feature="F14F18_30"
/feature="similarity to NLI-interacting factor isoform T1,
Gallus gallus, EMBL:AF189773"
/codon_start=1
/product="putative protein"
/protein_id="CAB87659.1"
/db_xref="GI:7573353"
/translation="MEDDSSPSRDLDAQNPYDRLALDTSTVDPNCNLDVSATYLA
MSSKRLCEVDREGDLSLTSVCMEDEDEDELDFDPYLFIKNLPNLSVVYPTRPVLL
PKOTRSCPTSLVLDLDELTVHSTLERCGGVDTFTPPYVNFEEHMYVVRKPLKRFM
EYVSRLEFIITFTASOSITAEQLNLVLDPRKLFRRHVRVRSQVFFDGNLKLDSVIG
RDLSTRVIVDNSPQAFQFQVENGVPLESWFNDSDEKLELLHLPFLSLIGVEDRPMI
AKKFNLRKIDAAVAAPEYPAEGDPEFR"
complement(16565..16786)
/feature="F14F18_30"
/number=1
complement(16787..17109)
/number=1
complement(17110..17447)
/feature="F14F18_30"
/number=2
complement(17448..17578)
```

Query Match 2.0%; Score 20; DB 13; Length 103638;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Mismatches 0; Indels 0; Caps 0;

QY 760 atttgaagaagattccaacc 779
|||||
DB 90372 ATTTGGAAGATTTCAGACC 90391

RESULT 22

HS173D1/c 117338 bp DNA PRI 23-NOV-1999
LOCUS Human DNA sequence from clone 173D1 on chromosome 1p36.21-36.33.
DEFINITION Contains ESTs, STSs and GSSs, complete sequence.
ACCESSION AL031984
VERSION AL031984.13 GI:4678432
KEYWORDS HTG; Cpg Island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 117338)
AUTHORS Moore, M.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT requests: clonerequests@sanger.ac.uk
On Apr 24, 1999 this sequence version replaced gi:4584747.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL
This sequence is the entire insert of clone 173D1. This sequence
has been finished according to sequencing map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
173D1 is from the library RPI1 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://pacpac.med.buffalo.edu/ VECTOR: pCIRpac2.
FEATURES
source
location/Qualifiers
1..117338
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p36.21-36.33"
/clone="RPI-173D1"
/clone_lib="RPI-1"
91..142
/note="13 copies 4 mer aggg 75% conserved"
repeat_region
585..992
/note="17 copies 24 mer 97% conserved"
repeat_region
1728..1903
/note="WER45 repeat: matches 1..178 of consensus"
repeat_region
1909..2078
/note="MIR repeat: matches 28..187 of consensus"
repeat_region
4058..4357
/note="AluY repeat: matches 1..298 of consensus"
repeat_region
4464..4571
/note="2 copies 54 mer 96% conserved"
repeat_region
4726..4844
/note="MIR repeat: matches 21..151 of consensus"
repeat_region
5494..5786
/note="AluX repeat: matches 1..292 of consensus"
repeat_region
6503..6807
/note="AluX repeat: matches 1..311 of consensus"
repeat_region
6809..6857
/note="MIR repeat: matches 101..150 of consensus"
repeat_region
6858..7181
/note="AluX repeat: matches 1..311 of consensus"
repeat_region
7182..7299
/note="MIR repeat: matches 150..262 of consensus"
repeat_region
8932..8993
/note="WER53 repeat: matches 126..189 of consensus"
repeat_region
8994..9305
/note="AluSg repeat: matches 1..311 of consensus"
repeat_region
9306..9440
/note="WER53 repeat: matches 1..126 of consensus"
repeat_region
10094..10654
/note="L2 repeat: matches 2152..2746 of consensus"
repeat_region
13515..13745
/note="L2 repeat: matches 1886..2084 of consensus"
repeat_region
13746..14051
/note="AluX repeat: matches 1..306 of consensus"
repeat_region
14052..14530
/note="L2 repeat: matches 2084..2680 of consensus"
repeat_region
14561..14871
/note="AluX repeat: matches 1..311 of consensus"
repeat_region
16630..16703
/note="MIR repeat: matches 101..176 of consensus"
misc_feature
complement(17036..17418)
/note="match: STS G11038"
repeat_region
19534..19676
/note="MIR repeat: matches 26..168 of consensus"
repeat_region
19811..20021
/note="L2 repeat: matches 2521..2705 of consensus"
repeat_region
20103..20402
/note="AluX repeat: matches 1..300 of consensus"
repeat_region
20551..20620
/note="L2 repeat: matches 2632..2701 of consensus"
repeat_region
20681..20878
/note="MIR repeat: matches 28..238 of consensus"
repeat_region
22445..22468
/note="L2 copies 2 mer ca 100% conserved"
repeat_region
22496..22714
/note="MIR repeat: matches 2..223 of consensus"
repeat_region
22965..23143
/note="MIR repeat: matches 71..232 of consensus"
repeat_region
23333..23460
/note="2 copies 64 mer 98% conserved"
repeat_region
24447..24757
/note="AluSg repeat: matches 1..309 of consensus"
repeat_region
25099..25400
/note="AluX repeat: matches 1..303 of consensus"
repeat_region
25523..25637
/note="Charlies repeat: matches 5..121 of consensus"
repeat_region
25767..26056
/note="AluX repeat: matches 1..291 of consensus"
repeat_region
26059..26270
/note="MIR repeat: matches 35..256 of consensus"
repeat_region
26313..26624
/note="AluY repeat: matches 1..312 of consensus"
repeat_region
27898..27939
/note="7 copies 6 mer tcttc 95% conserved"
repeat_region
29266..29305
/note="10 copies 4 mer aagg 80% conserved"
misc_feature
31618..31978
/note="match: EST A1350672"
repeat_region
33293..33452
/note="MIR repeat: matches 48..203 of consensus"
repeat_region
33832..34011
/note="90 copies 2 mer ac 71% conserved"
repeat_region
33832..33999
/note="7 copies 24 mer 73% conserved"
repeat_region
33832..34023
/note="3 copies 64 mer 72% conserved"
repeat_region
33832..33993
/note="3 copies 54 mer 74% conserved"
repeat_region
34018..34503
/note="9 copies 54 mer 67% conserved"
repeat_region
34018..34497
/note="20 copies 24 mer 66% conserved"
repeat_region
34018..34505
/note="122 copies 4 mer caca 65% conserved"
repeat_region
34018..34491


```

repeat_region /note="237 copies 2 mer ca 66% conserved"
34025..34504
repeat_region /note="48 copies 10 mer acacacacac 65% conserved"
34028..34507
repeat_region /note="80 copies 6 mer cacaca 65% conserved"
34065..34512
repeat_region /note="7 copies 64 mer 70% conserved"
34501..34914
repeat_region /note="3 copies 138 mer 70% conserved"
34508..34837
repeat_region /note="33 copies 10 mer acacacacac 66% conserved"
34510..34839
repeat_region /note="55 copies 6 mer acacac 66% conserved"
34511..34846
repeat_region /note="14 copies 24 mer 66% conserved"
34519..34842
repeat_region /note="6 copies 54 mer 67% conserved"
34519..34838
repeat_region /note="160 copies 2 mer ca 66% conserved"
35512..35541
misc_feature /note="15 copies 2 mer tt 90% conserved"
complement(36540..36839)
repeat_region /note="match: EST AA295567"
36962..37180
repeat_region /note="MIR repeat: matches 9..251 of consensus"
37142..37185
repeat_region /note="L2 repeat: matches 2655..2698 of consensus"
37989..38163
repeat_region /note="MIR repeat: matches 47..236 of consensus"
39487..39536
repeat_region /note="5 copies 10 mer gtgtgtgtgt 90% conserved"
39487..39528
repeat_region /note="21 copies 2 mer gt 98% conserved"
39487..39534
repeat_region /note="12 copies 4 mer gtgt 92% conserved"
39831..40133
repeat_region /note="Alusk repeat: matches 1..304 of consensus"
40195..40373
repeat_region /note="MIR repeat: matches 32..222 of consensus"
40574..40710
repeat_region /note="MER3 repeat: matches 9..149 of consensus"
40684..40748
repeat_region /note="Charlies repeat: matches 1..62 of consensus"
40795..41105
repeat_region /note="Alusk repeat: matches 1..309 of consensus"
complement(41047..41397)
misc_feature /note="match: STS L31008"
41244..41387
repeat_region /note="72 copies 2 mer tc 63% conserved"
41263..41538
repeat_region /note="46 copies 6 mer ttcttt 57% conserved"
41264..41527
repeat_region /note="11 copies 24 mer 61% conserved"
41271..41465
repeat_region /note="5 copies 39 mer 73% conserved"

Query Match 2.0%; Score 20; DB 92; Length 117338;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 caagattcaaacagctca 29
Db 68061 CAGATTCAAAACCCAGCTCA 68042

```

```

RESULT 23
AC027795/C AC027795 148508 bp DNA HTG 03-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-235B15 map 18, WORKING DRAFT
ACCESSION AC027795
VERSION AC027795.2 GI:7684508
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

```

SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 148508)
Birken, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE
Homo sapiens chromosome 18, clone RP11-235B15
JOURNAL
Unpublished
REFERENCE
AUTHORS
2 (bases 1 to 148508)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Girelli, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Menous, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Ollivar, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
TITLE
Direct Submission
JOURNAL
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 3, 2000 this sequence version replaced gi:7382637.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
COMMENT
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9034
Center clone name: 235_B_15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 138322 bases at least Q40
Consensus quality: 143906 bases at least Q30
Consensus quality: 145997 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 146908; sum-of-ctrls
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1262: contig of 1262 bp in length
* 1263 1362: gap of 100 bp
* 1363 1362: gap of 100 bp
* 1363 2466: contig of 1104 bp in length
* 2467 2566: gap of 100 bp
* 2567 3867: contig of 1301 bp in length
* 3868 3967: gap of 100 bp
* 3968 4372: contig of 405 bp in length
* 4373 4472: gap of 100 bp

```

```

* 4473 7936: contig of 3464 bp in length
* 7937 8036: gap of 100 bp
* 8037 11973: contig of 3937 bp in length
* 11974 12073: gap of 100 bp
* 12074 16804: contig of 4731 bp in length
* 16805 16904: gap of 100 bp
* 16905 22543: contig of 5639 bp in length
* 22544 22643: gap of 100 bp
* 22644 27921: contig of 5278 bp in length
* 27922 28021: gap of 100 bp
* 28022 33883: contig of 5862 bp in length
* 33884 33983: gap of 100 bp
* 33984 39792: contig of 5809 bp in length
* 39793 39892: gap of 100 bp
* 39893 49448: contig of 9556 bp in length
* 49449 49548: gap of 100 bp
* 49549 59577: contig of 10025 bp in length
* 59578 59677: gap of 100 bp
* 59678 69844: contig of 10167 bp in length
* 69845 69944: gap of 100 bp
* 69945 95609: contig of 25665 bp in length
* 95610 95709: gap of 100 bp
* 95710 120155: contig of 24446 bp in length
* 120156 120255: gap of 100 bp
* 120256 148508: contig of 28253 bp in length.
Location/Qualifiers
1. 148508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-235815"
/clone_lib="RPC1-11 Human Male BAC"
1. 1262
/note="assembly-fragment"
misc_feature 1363. 2466
/note="assembly-fragment"
misc_feature 2567. 3867
/note="assembly-fragment"
misc_feature 3968. 4372
/note="assembly-fragment"
clone_end:SP6
vector_side:right"
4473. 7936
/note="assembly-fragment"
misc_feature 8037. 11973
/note="assembly-fragment"
misc_feature 12074. 16804
/note="assembly-fragment"
clone_end:77
vector_side:right"
16905. 22543
/note="assembly-fragment"
misc_feature 22644. 27921
/note="assembly-fragment"
misc_feature 28022. 33883
/note="assembly-fragment"
misc_feature 33984. 39792
/note="assembly-fragment"
misc_feature 39893. 49448
/note="assembly-fragment"
misc_feature 49549. 59577
/note="assembly-fragment"
misc_feature 59678. 69844
/note="assembly-fragment"
misc_feature 69945. 95609
/note="assembly-fragment"
misc_feature 95710. 120155
/note="assembly-fragment"
misc_feature 120256. 148508
/note="assembly-fragment"
BASE COUNT 44364 a 29387 c 30187 g 42968 t 1602 others
ORIGIN

```

```

Query Match 2.0%; Score 20; DB 71; Length 148508;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 304 agaatacattccttgtaa 323
DB 9577 ACAATTCATCTTCCTTGGA 9558
|||||
RESULT 24
AC015525 149180 bp DNA HTG 12-MAR-2000
AC015525 Homo sapiens chromosome 15 clone RP11-110C15 map 15, WORKING DRAFT
LOCUS SEQUENCE, 9 unordered pieces.
DEFINITION
AC015525.3 GI:7108019
AC015525 HTG; HTGS_PHASE1; HTGS_DRAFT.
VERSION
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 149180)
Birken,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-110C15
Unpublished
2 (bases 1 to 149180)
Birken,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearlano,K., Dewar,K., Domingo,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,A.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliou,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.U., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:6553530.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4276
Center clone name: 110.C.15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 132714 bases at least Q40
Consensus quality: 142719 bases at least Q30
Consensus quality: 146304 bases at least Q20
Insert size: 156000; agarose-fp
Insert size: 148380; sum-of-ontigs
Quality coverage: 5.2 in Q20 bases; agarose-fp
Quality coverage: 5.5 in Q20 bases; sum-of-ontigs
-----

```

NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 24905: contig of 24905 bp in length
24906 25005: gap of unknown length
25006 38744: contig of 13739 bp in length
38745 38844: gap of unknown length
38845 53622: contig of 14778 bp in length
53623 53723: gap of unknown length
53723 65642: contig of 11920 bp in length
65643 73486: gap of unknown length
73487 73586: gap of unknown length
73587 81993: contig of 8307 bp in length
81994 88696: gap of unknown length
88697 97825: contig of 6703 bp in length
97826 97925: gap of unknown length
97926 103924: contig of 5999 bp in length
103925 104024: gap of unknown length
104025 109912: contig of 5888 bp in length
109913 110012: gap of unknown length
110013 116598: contig of 6586 bp in length
116599 120672: gap of unknown length
120673 120772: gap of unknown length
120773 125139: contig of 4367 bp in length
125140 125339: gap of unknown length
125340 128331: contig of 3092 bp in length
128332 128431: gap of unknown length
128432 131187: contig of 2756 bp in length
131188 131287: gap of unknown length
131288 134177: contig of 2890 bp in length
134178 134277: gap of unknown length
134278 137482: contig of 3205 bp in length
137483 137582: gap of unknown length
137583 138864: contig of 1282 bp in length
138865 138964: gap of unknown length
138965 142546: contig of 3582 bp in length
142547 142646: gap of unknown length
142647 145006: contig of 2360 bp in length
145007 145106: gap of unknown length
145107 147639: contig of 2533 bp in length
147640 147739: gap of unknown length
147740 149420: contig of 1681 bp in length
149421 149520: gap of unknown length
149521 150683: contig of 1163 bp in length.
```

FEATURES
SOURCE
Location/Qualifiers
1. 150683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-683J13"

BASE COUNT 40413 a 33700 c 32450 g 41878 t 2242 others
ORIGIN

Query Match 2.0%; Score 20; DB 74; Length 150683;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 449 gaccaggtgacaggtctc 468
|||||
Db 878 GACCAGGTGACAGGTCTC 859

RESULT 26
AC018876/c
LOCUS AC018876 165493 bp DNA HTG 07-JUL-2000

DEFINITION Homo sapiens chromosome 18 clone RP11-289E15, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
ACCESSION AC018876
VERSION AC018876.4 GI:7230959
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 165493)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 165493)
Waterston, R.H.
Direct Submission
Submitted (21-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:6850795.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Project Information
Center project name: H_NH0289E15
----- Summary Statistics -----
Sequencing vector: M13; 81%
Sequencing vector: plasmid; 19%
Chemistry: Dye-Primer ET; 81% of reads
Chemistry: Dye-terminator Big Dye; 19% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 157046 bases at least Q40
Consensus quality: 160223 bases at least Q30
Consensus quality: 161942 bases at least Q20
Insert size: 174000; agarose-fp
Quality coverage: 4.36 in Q20 bases; sum-of-contigs
Quality coverage: 4.61 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 1322: contig of 1322 bp in length
1323 1422: gap of unknown length
1423 8681: contig of 7259 bp in length
8682 8781: gap of unknown length
8782 28642: contig of 19861 bp in length
28643 28742: gap of unknown length
28743 46484: contig of 17742 bp in length
46485 46584: gap of unknown length
46585 68690: contig of 22106 bp in length
68691 68790: gap of unknown length
68791 99194: contig of 30404 bp in length
99195 99294: gap of unknown length
99295 131431: contig of 32137 bp in length
131432 131531: gap of unknown length
131532 165493: contig of 33962 bp in length.
```

FEATURES
source
Location/Qualifiers
1. 165493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"

BASE COUNT 52008 a 30658 c 30763 g 51358 t 706 others
ORIGIN

Query Match 2.0%; Score 20; DB 65; Length 165493;
Best Local Similarity 100.0%; Pred.No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 578 gcttcacaccttgctggtg 597
|||||
Db 11874 GCTTCACCCCTTGCTGTGG 11855

RESULT 27
AC021927
LOCUS Homo sapiens clone RP11-29P22, WORKING DRAFT SEQUENCE, 26 unordered
DEFINITION
AC021927 168997 bp DNA HTG 03-MAR-2000
AC021927.3 GI:7144973
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 168997)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome, clone RP11-29P22
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 168997)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
Chapel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
Deatellano, K., Dewar, K., Domingo, M., Doyle, M., Fennestor, J.,
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisanu, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:5984426.
All repeats were identified using RepeatMasker:
Smt. A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4856
Center clone name: 29_P_22
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156612 bases at least Q40
Consensus quality: 162417 bases at least Q30
Consensus quality: 164820 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 166497; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1046: contig of 1046 bp in length
1047 1146: gap of 100 bp
1147 2513: contig of 1367 bp in length
2514 2613: gap of 100 bp
2614 3854: contig of 1241 bp in length
3855 3954: gap of 100 bp
3955 5409: contig of 1455 bp in length
5410 5509: gap of 100 bp
5510 8344: contig of 2835 bp in length
8345 8444: gap of 100 bp
8445 11304: contig of 2860 bp in length
11305 11404: gap of 100 bp
11405 14151: contig of 2747 bp in length
14152 14251: gap of 100 bp
14252 17174: contig of 2923 bp in length
17175 17274: gap of 100 bp
17275 19939: contig of 2665 bp in length
19940 20039: gap of 100 bp
20040 23411: contig of 3372 bp in length
23412 23511: gap of 100 bp
23512 27130: contig of 3619 bp in length
27131 27230: gap of 100 bp
27231 29770: contig of 2540 bp in length
29771 29870: gap of 100 bp
29871 36121: contig of 6251 bp in length
36122 36221: gap of 100 bp
36222 41978: contig of 5757 bp in length
41979 42078: gap of 100 bp
42079 49159: contig of 7081 bp in length
49160 49259: gap of 100 bp
49260 57801: contig of 8542 bp in length
57802 57901: gap of 100 bp
57902 64360: contig of 6455 bp in length
64361 64460: gap of 100 bp
64461 72524: contig of 8064 bp in length
72525 72624: gap of 100 bp
72625 81728: contig of 9104 bp in length
81729 81828: gap of 100 bp
81829 92259: contig of 10431 bp in length
92260 92359: gap of 100 bp
92360 100902: contig of 8543 bp in length
100903 101002: gap of 100 bp
101003 110082: contig of 9080 bp in length
110083 110182: gap of 100 bp
110183 124956: contig of 14774 bp in length
124957 125056: gap of 100 bp
125057 139552: contig of 14456 bp in length
139553 139652: gap of 100 bp
139653 153065: contig of 13413 bp in length
153066 153165: gap of 100 bp
153166 168997: contig of 15832 bp in length.

FEATURES
source
1. 168997
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-29P22"
/clone_lib="RP11 Human Male BAC"
1. 1046
/note="assembly_fragment"
1147. 2513
/note="assembly_fragment"
2614. 3854
/note="assembly_fragment"
3955. 5409
/note="assembly_fragment"
5510. 8344
/note="assembly_fragment"

```

misc_feature      8445..11304
                  /note="assembly-fragment"
misc_feature      11405..14151
                  /note="assembly-fragment"
misc_feature      14252..17174
                  /note="assembly-fragment"
misc_feature      17275..19939
                  /note="assembly-fragment"
misc_feature      20040..23411
                  /note="assembly-fragment"
misc_feature      23512..27130
                  /note="assembly-fragment"
misc_feature      27231..29770
                  /note="assembly-fragment"
                  clone_end:77
misc_feature      29871..36121
                  /note="assembly-fragment"
misc_feature      36222..41978
                  /note="assembly-fragment"
misc_feature      42079..49159
                  /note="assembly-fragment"
misc_feature      49260..57801
                  /note="assembly-fragment"
misc_feature      57902..64360
                  /note="assembly-fragment"
misc_feature      64461..72524
                  /note="assembly-fragment"
misc_feature      72625..81728
                  /note="assembly-fragment"
misc_feature      81829..92259
                  /note="assembly-fragment"
misc_feature      92360..100902
                  /note="assembly-fragment"
                  clone_end:SP6
                  vector_side:right"
misc_feature      101003..110082
                  /note="assembly-fragment"
misc_feature      110183..124956
                  /note="assembly-fragment"
misc_feature      125057..139552
                  /note="assembly-fragment"
misc_feature      139653..153065
                  /note="assembly-fragment"
misc_feature      153166..168997
                  /note="assembly-fragment"
BASE COUNT      48830 a 34497 c 33807 g 49359 t 2504 others
ORIGIN

```

```

Query Match      2.0%: Score 20; DB 66; Length 168997;
Best Local Similarity 100.0%: Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 304 agaaatcattctcttgaaa 323
|||||
Db 110794 ACAATCATCTCTTGAAA 110813

```

```

RESULT 28
AC025191      177868 bp      DNA      HTG      03-MAY-2000
DEFINITION    Homo sapiens chromosome 4 clone RP11-285A15 map 4, WORKING DRAFT
SEQUENCE      43 unordered pieces.
ACCESSION     AC025191.2 GI:7684444
VERSION       AC025191.2
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 177868)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens chromosome 4, clone RP11-285A15
Unpublished
2 (bases 1 to 177868)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karakas,A.,
Klein,J., Lacroque,R., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., McDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKenna,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenaga,V., Morrow,J.,
Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE JOURNAL COMMENT

Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 3, 2000 this sequence version replaced gi:1188869.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB
Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 285.A.15

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 150228 bases at least Q40

Consensus quality: 163346 bases at least Q30

Consensus quality: 169409 bases at least Q20

Insert size: 185000; agarose-rip

Insert size: 173668; sum-of-contigs

Quality coverage: 2.9 in Q20 bases; agarose-fp

Quality coverage: 3.2 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1215: contig of 1215 bp in length
* 1216 1315: gap of 100 bp
* 1316 2523: contig of 1208 bp in length
* 2524 2623: gap of 100 bp
* 2624 3811: contig of 1188 bp in length
* 3812 3911: gap of 100 bp
* 3912 5309: contig of 1398 bp in length
* 5310 5409: gap of 100 bp
* 5410 7250: contig of 1841 bp in length
* 7251 7350: gap of 100 bp
* 7351 8970: contig of 1620 bp in length
* 8971 9070: gap of 100 bp
* 9071 10289: contig of 1219 bp in length
* 10290 10389: gap of 100 bp

```

```

* 10390 11577: contig of 1188 bp in length
* 11578 11677: gap of 100 bp
* 11678 12960: contig of 1283 bp in length
* 12961 13060: gap of 100 bp
* 13061 14891: contig of 1831 bp in length
* 14892 14991: gap of 100 bp
* 14992 16528: contig of 1537 bp in length
* 16529 16628: gap of 100 bp
* 16629 18711: contig of 2083 bp in length
* 18712 18811: gap of 100 bp
* 18812 20214: contig of 1403 bp in length
* 20215 20314: gap of 100 bp
* 20315 22282: contig of 1968 bp in length
* 22283 22382: gap of 100 bp
* 22383 24680: contig of 2298 bp in length
* 24681 24780: gap of 100 bp
* 24781 27243: contig of 2463 bp in length
* 27244 27343: gap of 100 bp
* 27344 29231: contig of 1888 bp in length
* 29232 29331: gap of 100 bp
* 29332 32345: contig of 3014 bp in length
* 32346 32445: gap of 100 bp
* 32446 35005: contig of 2560 bp in length
* 35006 35105: gap of 100 bp
* 35106 37773: contig of 2668 bp in length
* 37774 37873: gap of 100 bp
* 37874 40633: contig of 2760 bp in length
* 40634 40733: gap of 100 bp
* 40734 43133: contig of 2400 bp in length
* 43134 43233: gap of 100 bp
* 43234 45935: contig of 2702 bp in length
* 45936 46035: gap of 100 bp
* 46036 48895: contig of 2860 bp in length
* 48896 48995: gap of 100 bp
* 48996 51307: contig of 2312 bp in length
* 51308 51407: gap of 100 bp
* 51408 55313: contig of 3906 bp in length
* 55314 55413: gap of 100 bp
* 55414 58208: contig of 2795 bp in length
* 58209 58308: gap of 100 bp
* 58309 62616: contig of 4308 bp in length
* 62617 62716: gap of 100 bp
* 62717 67413: contig of 4697 bp in length
* 67414 67513: gap of 100 bp
* 67514 70733: contig of 3220 bp in length
* 70734 70833: gap of 100 bp
* 70834 75405: contig of 4572 bp in length
* 75406 75505: gap of 100 bp
* 75506 80523: contig of 5018 bp in length
* 80524 80623: gap of 100 bp
* 80624 86066: contig of 5443 bp in length
* 86067 86166: gap of 100 bp
* 86167 92000: contig of 5834 bp in length
* 92001 92100: gap of 100 bp
* 92101 96663: contig of 4563 bp in length
* 96664 96763: gap of 100 bp
* 96764 102888: contig of 6125 bp in length
* 102889 102988: gap of 100 bp
* 102989 110032: contig of 7044 bp in length
* 110033 110132: gap of 100 bp
* 110133 117903: contig of 7771 bp in length
* 117904 118003: gap of 100 bp
* 118004 125508: contig of 7505 bp in length
* 125509 125608: gap of 100 bp
* 125609 137315: contig of 11707 bp in length
* 137316 137415: gap of 100 bp
* 137416 148418: contig of 11003 bp in length
* 148419 148518: gap of 100 bp
* 148519 161620: contig of 13102 bp in length
* 161621 161720: gap of 100 bp
* 161721 177868: contig of 16148 bp in length.
Location/Qualifiers
1..177868
FEATURES
source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-285A15"
/clone_11b="RPC1-11 Human Male BAC"
1..1215
/note="assembly_fragment"
1316..2523
/note="assembly_fragment"
2624..3811
/note="assembly_fragment"
3912..5309
/note="assembly_fragment"
5410..7250
/note="assembly_fragment"
7351..8970
/note="assembly_fragment"
9071..10289
/note="assembly_fragment"
10390..11577
/note="assembly_fragment"
11678..12960
/note="assembly_fragment"
13061..14891
/note="assembly_fragment"
14992..16528
/note="assembly_fragment"
16629..18711
/note="assembly_fragment"
18812..20214
/note="assembly_fragment"
20315..22282
/note="assembly_fragment"
22383..24680
/note="assembly_fragment"
24781..27243
/note="assembly_fragment"
27344..29231
/note="assembly_fragment"
29332..32345
/note="assembly_fragment"

Query Match      2.0%: Score 20; DB 69; Length 177868;
Best Local Similarity 100.0%: Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      593 ggtggtgcccccaaccagc 612
Db 57383 ggtggtgcccccaaccagc 57402

RESULT 29
AP001899/c
LOCUS      AP001899
DEFINITION Homo sapiens chromosome 18 clone RP11-701C7 map 18q12, WORKING
DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION  AP001899
VERSION    AP001899.3 GI:9501833
KEYWORDS  HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE    Homo sapiens DNA, clone:RP11-701C7.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 183412)
REFERENCE  1 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 183,412 genomic DNA of 18q12
2 (bases 1 to 183412)
REFERENCE  2 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 183,412 genomic DNA of 18q12
AUTHORS
TITLE
Direct Submission

```

JOURNAL

Submitted (24-APR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail: hattori@gscc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923, Fax: 81-42-778-9924)
 On Jul 26, 2000 this sequence version replaced gi:8117550.

COMMENT

Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gscc.riken.go.jp
 Project Information
 Center project name: HumDraFl18
 Center clone name: Rp11-701C7
 Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 180400 bases at least Q40
 Consensus quality: 181718 bases at least Q30
 Consensus quality: 182304 bases at least Q20
 Insert size: 182712; sum-of-coverage
 Quality coverage: 10.87x in Q20 bases; sum-of-coverage
 NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 42740 contig of 42740 bp in length
 42841 79938 contig of 37098 bp in length
 80039 109211 contig of 29173 bp in length
 109312 135418 contig of 26107 bp in length
 135519 161816 contig of 26298 bp in length
 161917 174638 contig of 12722 bp in length
 174739 180277 contig of 5539 bp in length
 180378 183412 contig of 3035 bp in length.
 NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

FEATURES

source
 1. 183412
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q12"
 /clone="Rp11-701C7"
 1. 42740
 misc_feature
 /note="assembly_fragment"
 42841. 79938
 /note="assembly_fragment clone_end:SP6 vector_side:right"

misc_feature 80039. 109211
 /note="assembly_fragment clone_end:T7 vector_side:left"
 misc_feature 109312. 135418
 /note="assembly_fragment"
 misc_feature 135519. 161816
 /note="assembly_fragment"
 misc_feature 161917. 174638
 /note="assembly_fragment"
 misc_feature 174739. 180277
 /note="assembly_fragment"
 misc_feature 180378. 183412
 /note="assembly_fragment"
 BASE COUNT 49688 a 39575 c 41016 g 52433 t 700 others
 ORIGIN

Query Match 2.0%; Score 20; DB 83; Length 183412;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 211 ggtgagcttccacttc 230
 Db 121048 GGTGAGCTTCCACTTC 121029

RESULT 30

AC009591

LOCUS

DEFINITION

AC009591

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCES

AUTHORS

2 (bases 1 to 183855)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,D., Barna,N., Beckwith,R., Benn,J., Brown,A., Castle,A., Cerny,J., Collange,M., Collins,S., Collins,A., Cooke,P., Dearlano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lechoczy,J., Lieu,C., Locke,K., MacDonald,P., Marguis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Moll,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Nijhoff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,J.A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Toruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
 Direct Submission
 Submitted (27-APR-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 28, 2000 this sequence version replaced gi:6094611.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: I1744
 Center clone name: 335_K_21
 Summary Statistics


```

Sequencing vector: M13, M77815; 100% of reads
Chemistry: Dye-Primer-amer sham; 2% of reads
Chemistry: Dye-terminator Big Dye; 98% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158308 bases at least Q40
Consensus quality: 170972 bases at least Q30
Consensus quality: 177434 bases at least Q20
Insert size: 190000; agarose-1p
Insert size: 183355; sum-of-contigs
Quality coverage: 5.2 in Q20 bases; agarose-1p
Quality coverage: 5.4 in Q20 bas.
NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1031: contig of 1031 bp in length
* 1032 1131: gap of 100 bp
* 1132 2356: contig of 1225 bp in length
* 2357 2456: gap of 100 bp
* 2457 3536: contig of 1080 bp in length
* 3537 3636: gap of 100 bp
* 3637 8344: contig of 4708 bp in length
* 8345 8444: gap of 100 bp
* 8445 45297: contig of 36853 bp in length
* 45298 45397: gap of 100 bp
* 45398 183855: contig of 138458 bp in length.
FEATURES
Source
1. 183855
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-335K21"
/clone_11b="RP11-11 Human Male BAC"
1. 1031
/note="assembly_fragment"
1132. 2356
/note="assembly_fragment"
2457. 3536
/note="assembly_fragment"
3637. 8344
/note="assembly_fragment"
8445. 45297
/note="assembly_fragment"
clone_end:r7
vector_side:right"
45398. 183855
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
BASE COUNT 62884 a 32653 c 31457 g 56345 t 516 others
ORIGIN
Query Match 2.0%; Score 20; DB 61; Length 183855;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY 593 ggtgtgtgccccaccaccagc 612
|||||
Db 61148 GGTGTGTGCCCCCAACACGAC 61167
RESULT 31
AC022446 184855 bp DNA HTG 06-MAY-2000
LOCUS Homo sapiens chromosome 5 clone RP11-5H13, WORKING DRAFT SEQUENCE,
10 unordered pieces.
ACCESSION AC022446

```

```

VERSION      AC022446.3  GI:7711748
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        DOE Joint Genome Institute.
REFERENCE    DOE Joint Genome Institute.
AUTHORS      Sequencing of Human Chromosome 5
TITLE        2 (bases 1 to 18485)
REFERENCE    DOE Joint Genome Institute.
AUTHORS      DOE Joint Genome Institute.
TITLE        Direct Submission
REFERENCE    Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint
AUTHORS      Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
TITLE        On May 6, 2000 this sequence version replaced gi:7341552.
COMMENT      -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 408466
            Center clone name: RPCI-11_5H13
            -----
            Summary Statistics
            Consensus quality: 173751 bases at least Q40
            Consensus quality: 180112 bases at least Q30
            Consensus quality: 181210 bases at least Q20
            Estimated insert size: 194880; agarose-1p estimation
            Estimated insert size: 183955; sum-of-contrigs estimation
            Quality coverage: 5.29 in Q20 bases; agarose-1p estimation
            Quality coverage: 5.6 in Q20 bases; sum-of-contrigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contrigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contrigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            1
            * 4597: contrig of 4597 bp in length
            * 4598
            * 4697: gap of unknown length
            * 4698
            * 11489: contrig of 6792 bp in length
            * 11490
            * 11589: gap of unknown length
            * 11590
            * 16331: contrig of 4742 bp in length
            * 16332
            * 16431: gap of unknown length
            * 16432
            * 24922: contrig of 8491 bp in length
            * 24923
            * 25022: gap of unknown length
            * 25023
            * 38500: contrig of 13478 bp in length
            * 38501
            * 38600: gap of unknown length
            * 38601
            * 57517: contrig of 18917 bp in length
            * 57518
            * 57617: gap of unknown length
            * 57618
            * 78268: contrig of 20651 bp in length
            * 78369
            * 78368: gap of unknown length
            * 103425: contrig of 25057 bp in length
            * 103426
            * 103525: gap of unknown length
            * 103526
            * 135753: contrig of 32228 bp in length
            * 135754
            * 135853: gap of unknown length
            * 135854
            * 184855: contrig of 49002 bp in length.
            Location/Qualifiers
            1. 184855
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="RP11-5H13"
            /clone_1ib="RPCI human BAC library 11"
            BASE COUNT  55073 a 36889 c 37015 g 54952 t          926 others
            ORIGIN
Query Match      2.0%; Score 20; DB 67; Length 184855;
Best Local Similarity 100.0%; Pred. No. 7.9;

```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 271 cccagcccccagagagaaaa 290
 ||||||||||||||||
 Db 182417 CCCAGCCCCAGCAGAGAAAA 182436

RESULT 32
 AC090095
 LOCUS AC090095
 DEFINITION Homo sapiens chromosome 8 clone RP11-296C13 map 8, WORKING DRAFT
 AC090095
 AC090095.2 GI:13357359
 HTG: HTGS_PHASE1, HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 192059)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 JOURNAL Homo sapiens chromosome 8, clone RP11-296C13
 REFERENCE 2 (bases 1 to 192059)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Batra,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
 Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collamore,A., Cooke,P., Dekrellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
 Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Karas,A., Lacroque,K., Lamazares,R., Landers,T.,
 Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
 Margulis,N., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
 McHeeters,R., Meldrim,J., Menus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nordu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
 Ribback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M.,
 Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
 Sougnaz,C., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
 Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

TITLE
 JOURNAL Submitted (14-FEB-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 16, 2001 this sequence version replaced gi:12830252.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L12760

Center clone name: 296_C_13

Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 183542 bases at least Q40
 Consensus quality: 187455 bases at least Q40
 Consensus quality: 188948 bases at least Q30
 Insert size: 189859; sum-of-contigs
 Quality coverage: 4.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 957: contig of 957 bp in length
 * 958 1057: gap of 100 bp
 * 1058 2239: contig of 1182 bp in length
 * 2240 2339: gap of 100 bp
 * 2340 3382: contig of 1043 bp in length
 * 3383 3482: gap of 100 bp
 * 3483 4503: contig of 1021 bp in length
 * 4504 4603: gap of 100 bp
 * 4604 7524: contig of 2921 bp in length
 * 7525 7624: gap of 100 bp
 * 7625 9358: contig of 1734 bp in length
 * 9359 9458: gap of 100 bp
 * 9459 12163: contig of 2705 bp in length
 * 12164 12263: gap of 100 bp
 * 12264 16386: contig of 4123 bp in length
 * 16387 16486: gap of 100 bp
 * 16487 19796: contig of 3310 bp in length
 * 19797 19896: gap of 100 bp
 * 19897 22758: contig of 2862 bp in length
 * 22759 22858: gap of 100 bp
 * 22859 26552: contig of 3694 bp in length
 * 26553 26652: gap of 100 bp
 * 26653 28501: contig of 1849 bp in length
 * 28502 28601: gap of 100 bp
 * 28602 32930: contig of 4329 bp in length
 * 32931 33030: gap of 100 bp
 * 33031 33985: contig of 6555 bp in length
 * 33986 39685: gap of 100 bp
 * 39686 46281: contig of 6596 bp in length
 * 46282 46381: gap of 100 bp
 * 46382 52654: contig of 6273 bp in length
 * 52655 52754: gap of 100 bp
 * 52755 57436: contig of 4682 bp in length
 * 57437 57536: gap of 100 bp
 * 57537 64278: contig of 6742 bp in length
 * 64279 64378: gap of 100 bp
 * 64379 73637: contig of 9259 bp in length
 * 73638 73737: gap of 100 bp
 * 73738 87928: contig of 14191 bp in length
 * 87929 88028: gap of 100 bp
 * 88029 106571: contig of 18543 bp in length
 * 106572 106671: gap of 100 bp
 * 106672 158839: contig of 52168 bp in length
 * 158840 158939: gap of 100 bp
 * 158940 192059: contig of 33120 bp in length.

FEATURES

source
 1..192059
 /organism="Homo sapiens"
 /db_xref="taxon:3606"
 /chromosome="8"
 /map="8"
 /clone="RP11-296C13"
 /clone.lib="RPC1-11 Human Male BAC"
 1..957
 /note="assembly-fragment
 clone_end:SP6
 vector_side:left"
 1058..2239
 /note="assembly-fragment"
 2340..3382
 /note="assembly-fragment"
 3483..4503
 /note="assembly-fragment"
 4604..7524
 /note="assembly-fragment"
 7625..9358
 /note="assembly-fragment"
 9459..12163
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature

* 45192 50472: contig of 5281 bp in length
* 50473 50572: gap of 100 bp
* 50573 55996: contig of 5424 bp in length
* 55997 56096: gap of 100 bp
* 56097 62551: contig of 6455 bp in length
* 62552 62651: gap of 100 bp
* 62652 67720: contig of 5069 bp in length
* 67721 67820: gap of 100 bp
* 67821 74113: contig of 6293 bp in length
* 74114 74213: gap of 100 bp
* 74214 80371: contig of 6158 bp in length
* 80372 80471: gap of 100 bp
* 80472 85618: contig of 5147 bp in length
* 85619 85718: gap of 100 bp
* 85719 96060: contig of 10342 bp in length
* 96061 96160: gap of 100 bp
* 96161 103014: contig of 6854 bp in length
* 103015 103114: gap of 100 bp
* 103115 114047: contig of 10933 bp in length
* 114048 114147: gap of 100 bp
* 114148 125353: contig of 11206 bp in length
* 125354 125453: gap of 100 bp
* 125454 137602: contig of 12149 bp in length
* 137603 137702: gap of 100 bp
* 137703 150975: contig of 13273 bp in length
* 150976 151075: gap of 100 bp
* 151076 173481: contig of 22406 bp in length
* 173482 173581: gap of 100 bp
* 173582 195383: contig of 21802 bp in length.

FEATURES

Source

1. 195383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-71F23"
/clone_11b="RP11 Human Male BAC"
1. 1875
/note="assembly-fragment"
misc_feature
1976..3161
/note="assembly-fragment"
misc_feature
3262..4663
/note="assembly-fragment"
misc_feature
4764..6577
/note="assembly-fragment"
misc_feature
6678..8647
/note="assembly-fragment"
misc_feature
8748..10746
/note="assembly-fragment"
misc_feature
10847..12332
/note="assembly-fragment"
misc_feature
12433..15508
/note="assembly-fragment"
misc_feature
15609..18286
/note="assembly-fragment"
misc_feature
18387..20319
/note="assembly-fragment"
clone_end:r7
vector_side:right
20420..23111
/note="assembly-fragment"
misc_feature
23212..26610
/note="assembly-fragment"
misc_feature
26711..30644
/note="assembly-fragment"
misc_feature
30745..36084
/note="assembly-fragment"
misc_feature
36185..39659
/note="assembly-fragment"
misc_feature
39760..45091
/note="assembly-fragment"
misc_feature
45192..50472
/note="assembly-fragment"

misc_feature 50573..55996
/note="assembly-fragment"
misc_feature 56097..62551
/note="assembly-fragment"
misc_feature 62652..67720
/note="assembly-fragment"
misc_feature 67821..74113
/note="assembly-fragment"
misc_feature 74214..80371
/note="assembly-fragment"
misc_feature 80472..85618
/note="assembly-fragment"
misc_feature 85719..96060
/note="assembly-fragment"
misc_feature 96161..103014
/note="assembly-fragment"
clone_end:SP6
vector_side:left
103115..114047
/note="assembly-fragment"
misc_feature 114148..125353
/note="assembly-fragment"
misc_feature 125454..137602
/note="assembly-fragment"
misc_feature 137703..150975
/note="assembly-fragment"
misc_feature 151076..173481
/note="assembly-fragment"
misc_feature 173582..195383
/note="assembly-fragment"

BASE COUNT 52065 a 43628 c 42262 g 54407 t 3021 others

Query Match 2.0%; Score 20; DB 62; Length 195383;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 211 gctgcagctcccaacttc 230
|||||
Db 25633 GCTGCAGCTCCCACTTCC 25652

RESULT 34
AL358492 197788 bp DNA HTG 09-MAR-2001
LOCUS AL358492/C
DEFINITION Homo sapiens chromosome 1 clone RP11-340B24, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AL358492 GI:13274738
VERSION AL358492.9 GI:13274738
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 197788)
AUTHORS Slims, S.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Mar 12, 2001 this sequence version replaced gi:13273744.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba340B24
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 194817 bases at least Q40
Consensus quality: 195716 bases at least Q30

COMMENT
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba340B24
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 194817 bases at least Q40
Consensus quality: 195716 bases at least Q30

Consensus quality: 196299 bases at least Q20
Insert size: 196988; sum-of-contigs
Insert size: 185634; 10.0% error; agarose-fp
Quality coverage: 7.00x in Q20 bases; sum-of-contigs quality
coverage: 7.67x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 29856: contig of 29856 bp in length
29857 29956: gap of 100 bp
29957 56368: contig of 26412 bp in length
56369 56468: gap of 100 bp
56469 66708: contig of 10240 bp in length
66709 66808: gap of 100 bp
66809 84676: contig of 17868 bp in length
84677 84776: gap of 100 bp
84777 121388: contig of 36612 bp in length
121389 121488: gap of 100 bp
121489 128913: contig of 7425 bp in length
128914 129013: gap of 100 bp
129014 162549: contig of 33536 bp in length
162550 162649: gap of 100 bp
162650 185166: contig of 22517 bp in length
185167 185266: gap of 100 bp
185267 197788: contig of 12522 bp in length.

FEATURES

Location/Qualifiers

1..197788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-340B24"
/clone_lib="RPCT-11.2"
1..29856
/note="assembly_fragment:00892
fragment_chain:1
clone_end:SP6
vector_side:left"
misc_feature
29957..56368
/note="assembly_fragment:02882
fragment_chain:1"
56469..66708
/note="assembly_fragment:01981
fragment_chain:2"
66809..84676
/note="assembly_fragment:02337
fragment_chain:2"
84777..121388
/note="assembly_fragment:00717
fragment_chain:2"
121489..128913
/note="assembly_fragment:03116
fragment_chain:2"
129014..162549
/note="assembly_fragment:00061
fragment_chain:2"
162650..185166
/note="assembly_fragment:01279
fragment_chain:2"
185267..197788
/note="assembly_fragment:00238
fragment_chain:2"
BASE COUNT 45671 a 51501 c 50781 g 49025 t 810 others
ORIGIN

Query Match 2.0%; Score 20; DB 80; Length 197788;
Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 caagatccaaccgactca 29

Db 15096 CAAGATTCAACCACGACTCA 15077

RESULT 35

AC018443

LOCUS

DEFINITION

AC018443

ACCESSION

AC018443

VERSION

AC018443.6

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 199812)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,

Boguslavsky,L., Boucknight,B., Brown,A., Castle,A., Colangelo,M.,

Collins,S., Collymore,A., Cooke,P., Dearlano,K., Dewar,K.,

Domino,M., Doyle,M., Fenebor,J., Ferreira,P., FitzHugh,W.,

Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B.,

Heard,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L.,

Karatas,A., Klein,J., Landers,T., Lechoczky,J., Lien,C., Locke,K.,

Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T.,

O'Donnell,P., Peterson,K., Pierre,N., Pollara,V., Riley,R.,

Rothman,D., Roy,A., Santos,R., Severy,P., Stange-Thomann,N.,

Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodores,J.,

Tirrell,A., Vassiliou,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,

Zimmer,A., and Zody,M.

Direct Submission

Submitted (10-DEC-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 25, 2001 this sequence version replaced gi:1249139.

All repeats were identified using RepeatMasker:

Smit,A.F.A. & Green,P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: 15117

Center clone name: 674.A.24

----- Summary Statistics

Sequencing vector: M13; M77815; 63% of reads

Sequencing vector: Plasmid; n/a; 37% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 198051 bases at least Q40

Consensus quality: 198846 bases at least Q30

Consensus quality: 199187 bases at least Q20

Insert size: 194000; agarose-fp

Insert size: 199412; sum-of-contigs

Quality coverage: 7.7 in Q20 bases; agarose-fp

Quality coverage: 7.5 in Q20 b.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 5 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

```

* be preserved.
* 1 15366: contig of 15366 bp in length
* 15367 15466: gap of 100 bp
* 15467 36444: contig of 20978 bp in length
* 36445 36544: gap of 100 bp
* 36545 107322: contig of 70778 bp in length
* 107323 107422: gap of 100 bp
* 107423 162742: contig of 55320 bp in length
* 162743 162842: gap of 100 bp
* 162843 199812: contig of 36970 bp in length.
Location/Qualifiers
1. 199812
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-674A24"
/clone_lib="RPCT-11 Human Male BAC"
misc_feature
1. 15366
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
15467..36444
/note="assembly_fragment"
misc_feature
36545..107322
/note="assembly_fragment"
misc_feature
107423..162742
/note="assembly_fragment"
misc_feature
162843..199812
/note="assembly_fragment"
clone_end:r7
vector_side:right"
BASE COUNT 66339 a 34885 c 34539 g 63649 t 400 others
ORIGIN
Query Match 2.0%; Score 20; DB 65; Length 199812;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 888 atggggaagatgccacct 907
|||||
Db 87641 ATGGGGAAGATGCCACCT 87660
|||||

RESULT 36
AC068379 222876 bp DNA HTG 22-OCT-2000
LOCUS AC068379 Homo sapiens chromosome 3 clone RP11-537116 map 3, LOW-PASS
DEFINITION AC068379
SEQUENCE SAMPLING.
AC068379 4 GI:10945763
VERSION AC068379.4
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 222876)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 3, clone RP11-537116
TITLE Unpublished
2 (bases 1 to 222876)
AUTHORS Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouhagalter,B., Brown,M., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glende,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,

```

```

TITLE JOURNAL
COMMENT
* NOTE: This record contains 278 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 728: contig of 728 bp in length
* 729 828: gap of 100 bp
* 829 1559: contig of 731 bp in length
* 1560 1659: gap of 100 bp
* 1660 2382: contig of 723 bp in length
* 2383 2482: gap of 100 bp
* 2483 3231: contig of 749 bp in length
* 3232 3331: gap of 100 bp
* 3332 4056: contig of 725 bp in length
* 4057 4156: gap of 100 bp
* 4157 4881: contig of 725 bp in length
* 4882 4981: gap of 100 bp
* 4982 5714: contig of 733 bp in length
* 5715 5814: gap of 100 bp
* 5815 6526: contig of 712 bp in length
* 6527 6626: gap of 100 bp
* 6627 7343: contig of 717 bp in length
* 7344 7443: gap of 100 bp
* 7444 8173: contig of 730 bp in length
* 8174 8273: gap of 100 bp
* 8274 8980: contig of 707 bp in length
* 8981 9080: gap of 100 bp
* 9081 9788: contig of 708 bp in length
* 9789 9888: gap of 100 bp
* 9889 10620: contig of 732 bp in length
* 10621 10720: gap of 100 bp
* 10721 11471: contig of 751 bp in length
* 11472 11571: gap of 100 bp
* 11572 12317: contig of 746 bp in length
* 12318 12417: gap of 100 bp
* 12418 13148: contig of 731 bp in length
* 13149 13248: gap of 100 bp
* 13249 13973: contig of 725 bp in length
* 13974 14073: gap of 100 bp
* 14074 14812: contig of 739 bp in length
* 14813 14912: gap of 100 bp
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
Center project name: 537-I_16
Center clone name: 537-I_16
-----
Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 22, 2000 this sequence version replaced g1:10305231.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
Center project name: 537-I_16
Center clone name: 537-I_16
-----

```

```

* 14913 15633: contig of 721 bp in length
* 15634 15733: gap of 100 bp
* 15734 16475: contig of 742 bp in length
* 16476 16575: gap of 100 bp
* 16576 17284: contig of 709 bp in length
* 17285 17384: gap of 100 bp
* 17385 18078: contig of 694 bp in length
* 18079 18178: gap of 100 bp
* 18179 18888: contig of 710 bp in length
* 18889 18988: gap of 100 bp
* 18989 19732: contig of 744 bp in length
* 19733 19832: gap of 100 bp
* 19833 20559: contig of 727 bp in length
* 20560 20659: gap of 100 bp
* 20660 21391: contig of 732 bp in length
* 21392 21491: gap of 100 bp
* 21492 22218: contig of 727 bp in length
* 22219 22318: gap of 100 bp
* 22319 23037: contig of 719 bp in length
* 23038 23137: gap of 100 bp
* 23138 23846: contig of 709 bp in length
* 23847 23946: gap of 100 bp
* 23947 24668: contig of 722 bp in length
* 24669 24768: gap of 100 bp
* 24769 25500: contig of 732 bp in length
* 25501 25600: gap of 100 bp
* 25601 26321: contig of 721 bp in length
* 26322 26421: gap of 100 bp
* 26422 27131: contig of 710 bp in length
* 27132 27231: gap of 100 bp
* 27232 27952: contig of 721 bp in length
* 27953 28052: gap of 100 bp
* 28053 28736: contig of 684 bp in length
* 28737 28836: gap of 100 bp
* 28837 29582: contig of 746 bp in length
* 29583 29682: gap of 100 bp
* 29683 30397: contig of 715 bp in length
* 30398 30497: gap of 100 bp
* 30498 31242: contig of 745 bp in length
* 31243 31342: gap of 100 bp
* 31343 32069: contig of 727 bp in length
* 32070 32169: gap of 100 bp
* 32170 32889: contig of 720 bp in length
* 32890 32989: gap of 100 bp
* 32990 33723: contig of 734 bp in length
* 33724 33823: gap of 100 bp
* 33824 34523: contig of 700 bp in length
* 34524 34623: gap of 100 bp
* 34624 35342: contig of 719 bp in length
* 35343 35442: gap of 100 bp
* 35443 36172: contig of 730 bp in length
* 36173 36272: gap of 100 bp
* 36273 36988: contig of 716 bp in length
* 36989 37088: gap of 100 bp
* 37089 37820: contig of 732 bp in length
* 37821 37920: gap of 100 bp
* 37921 38636: contig of 716 bp in length
* 38637 38736: gap of 100 bp
* 38737 39455: contig of 719 bp in length
* 39456 39555: gap of 100 bp
* 39556 40249: contig of 694 bp in length
* 40250 40349: gap of 100 bp
* 40350 41074: contig of 725 bp in length
* 41075 41174: gap of 100 bp
* 41175 41907: contig of 733 bp in length
* 41908 42007: gap of 100 bp
* 42008 42726: contig of 719 bp in length
* 42727 42826: gap of 100 bp
* 42827 43579: contig of 753 bp in length
* 43580 43679: gap of 100 bp
* 43680 44374: contig of 695 bp in length
* 44375 44474: gap of 100 bp
* 44475 45180: contig of 706 bp in length

```

```

* 45181 45280: gap of 100 bp
* 45281 45979: contig of 699 bp in length
* 45980 46079: gap of 100 bp
* 46080 46821: contig of 742 bp in length
* 46822 46921: gap of 100 bp
* 46922 47645: contig of 724 bp in length
* 47646 47745: gap of 100 bp
* 47746 48465: contig of 720 bp in length
* 48466 48565: gap of 100 bp
* 48566 49289: contig of 724 bp in length
* 49290 49389: gap of 100 bp
* 49390 50107: contig of 718 bp in length
* 50108 50207: gap of 100 bp
* 50208 50925: contig of 718 bp in length
* 50926 51025: gap of 100 bp
* 51026 51736: contig of 711 bp in length
* 51737 51836: gap of 100 bp
* 51837 52543: contig of 707 bp in length
* 52544 52643: gap of 100 bp
* 52644 53346: contig of 703 bp in length
* 53347 53446: gap of 100 bp
* 53447 54192: contig of 746 bp in length
* 54193 54292: gap of 100 bp
* 54293 55014: contig of 722 bp in length
* 55015 55114: gap of 100 bp
* 55115 55824: contig of 710 bp in length
* 55825 55924: gap of 100 bp
* 55925 56651: contig of 727 bp in length

Query Match      2.0%  Score 20; DB 73; Length 222876;
Best Local Similarity 100.00%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 gaccagtgagcaggtcct 468
Db 106725 GACCAGTGTGACAGGTCTT 106744
|||||
|||||

RESULT 37
AC068379/c      222876 bp      DNA      22-OCT-2000
LOCUS
DEFINITION
Homo sapiens chromosome 3 clone RP11-537116 map 3, LOW-PASS
SEQUENCE SAMPLING.
AC068379
AC068379.4 GI:10945763
VERSION
HTG: HTGS_PHASE0.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 222876)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 3, clone RP11-537116
Unpublished
2 (bases 1 to 222876)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,T., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Meldrum,J., Menus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

```

JOURNAL
COMMENT

Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 22, 2000 this sequence version replaced gi:10305231.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L10217

Center clone name: 537_L16

* NOTE: This record contains 278 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 728: contig of 728 bp in length
* 729 828: gap of 100 bp
* 829 1559: contig of 731 bp in length
* 1560 1659: gap of 100 bp
* 1660 2382: contig of 723 bp in length
* 2383 2482: gap of 100 bp
* 2483 3231: contig of 749 bp in length
* 3232 3331: gap of 100 bp
* 3332 4056: contig of 725 bp in length
* 4057 4156: gap of 100 bp
* 4157 4881: contig of 725 bp in length
* 4882 4981: gap of 100 bp
* 4982 5714: contig of 733 bp in length
* 5715 5814: gap of 100 bp
* 5815 6526: contig of 712 bp in length
* 6527 6626: gap of 100 bp
* 6627 7343: contig of 717 bp in length
* 7344 7443: gap of 100 bp
* 7444 8173: contig of 730 bp in length
* 8174 8273: gap of 100 bp
* 8274 8980: contig of 707 bp in length
* 8981 9080: gap of 100 bp
* 9081 9788: contig of 708 bp in length
* 9789 9888: gap of 100 bp
* 9889 10620: contig of 732 bp in length
* 10621 10720: gap of 100 bp
* 10721 11471: contig of 751 bp in length
* 11472 11571: gap of 100 bp
* 11572 12317: contig of 746 bp in length
* 12318 12417: gap of 100 bp
* 12418 13148: contig of 731 bp in length
* 13149 13248: gap of 100 bp
* 13249 13973: contig of 725 bp in length
* 13974 14073: gap of 100 bp
* 14074 14812: contig of 739 bp in length
* 14813 14912: gap of 100 bp
* 14913 15633: contig of 721 bp in length
* 15634 15733: gap of 100 bp
* 15734 16475: contig of 742 bp in length
* 16476 16575: gap of 100 bp
* 16576 17284: contig of 709 bp in length
* 17285 17384: gap of 100 bp
* 17385 18078: contig of 694 bp in length

* 18079 18178: gap of 100 bp
* 18179 18888: contig of 710 bp in length
* 18889 18988: gap of 100 bp
* 18989 19732: contig of 744 bp in length
* 19733 19832: gap of 100 bp
* 19833 20559: contig of 727 bp in length
* 20560 20659: gap of 100 bp
* 20660 21391: contig of 732 bp in length
* 21392 21491: gap of 100 bp
* 21492 22218: contig of 727 bp in length
* 22219 22318: gap of 100 bp
* 22319 23037: contig of 719 bp in length
* 23038 23137: gap of 100 bp
* 23138 23846: contig of 709 bp in length
* 23847 23946: gap of 100 bp
* 23947 24668: contig of 722 bp in length
* 24669 24768: gap of 100 bp
* 24769 25500: contig of 732 bp in length
* 25501 25600: gap of 100 bp
* 25601 26321: contig of 721 bp in length
* 26322 26421: gap of 100 bp
* 26422 27131: contig of 710 bp in length
* 27132 27231: gap of 100 bp
* 27232 27952: contig of 721 bp in length
* 27953 28052: gap of 100 bp
* 28053 28736: contig of 684 bp in length
* 28737 28836: gap of 100 bp
* 28837 29582: contig of 746 bp in length
* 29583 29682: gap of 100 bp
* 29683 30397: contig of 715 bp in length
* 30398 30497: gap of 100 bp
* 30498 31242: contig of 745 bp in length
* 31243 31342: gap of 100 bp
* 31343 32069: contig of 727 bp in length
* 32070 32169: gap of 100 bp
* 32170 32889: contig of 720 bp in length
* 32890 32989: gap of 100 bp
* 32990 33723: contig of 734 bp in length
* 33724 33823: gap of 100 bp
* 33824 34523: contig of 700 bp in length
* 34524 34623: gap of 100 bp
* 34624 35342: contig of 719 bp in length
* 35343 35442: gap of 100 bp
* 35443 36172: contig of 730 bp in length
* 36173 36272: gap of 100 bp
* 36273 36988: contig of 716 bp in length
* 36989 37088: gap of 100 bp
* 37089 37820: contig of 732 bp in length
* 37821 37920: gap of 100 bp
* 37921 38636: contig of 716 bp in length
* 38637 38736: gap of 100 bp
* 38737 39455: contig of 719 bp in length
* 39456 39555: gap of 100 bp
* 39556 40249: contig of 694 bp in length
* 40250 40349: gap of 100 bp
* 40350 41074: contig of 725 bp in length
* 41075 41174: gap of 100 bp
* 41175 41907: contig of 733 bp in length
* 41908 42007: gap of 100 bp
* 42008 42726: contig of 719 bp in length
* 42727 42826: gap of 100 bp
* 42827 43579: contig of 753 bp in length
* 43580 43679: gap of 100 bp
* 43680 44374: contig of 695 bp in length
* 44375 44474: gap of 100 bp
* 44475 45180: contig of 706 bp in length
* 45181 45280: gap of 100 bp
* 45281 45979: contig of 699 bp in length
* 45980 46079: gap of 100 bp
* 46080 46821: contig of 742 bp in length
* 46822 46921: gap of 100 bp
* 46922 47645: contig of 724 bp in length
* 47646 47745: gap of 100 bp


```

* 47746 48465: contig of 720 bp in length
* 48466 48565: gap of 100 bp
* 48566 49289: contig of 724 bp in length
* 49290 49389: gap of 100 bp
* 49390 50107: contig of 718 bp in length
* 50108 50207: gap of 100 bp
* 50208 50925: contig of 718 bp in length
* 50926 51025: gap of 100 bp
* 51026 51736: contig of 711 bp in length
* 51737 51836: gap of 100 bp
* 51837 52543: contig of 707 bp in length
* 52544 52643: gap of 100 bp
* 52644 53346: contig of 703 bp in length
* 53347 53446: gap of 100 bp
* 53447 54192: contig of 746 bp in length
* 54193 54292: gap of 100 bp
* 54293 55014: contig of 722 bp in length
* 55015 55114: gap of 100 bp
* 55115 55824: contig of 710 bp in length
* 55825 55924: gap of 100 bp
* 55925 56651: contig of 727 bp in length

```

Query Match 2.0%; Score 20; DB 73; Length 222876;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 449 gaccagtgacaggtcct 468
 DB 222603 GACCAGTGTGACAGGTCT 222584

```

RESULT 38
LOCUS AC004615/c 235141 bp DNA PRI 03-FEB-2000
DEFINITION Homo sapiens BAC clone GSI-405L21 from 5p15.2, complete sequence.
ACCESSION AC004615
VERSION AC004615.1 GI:3080660
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 235141)
Kailicki, J. and Harmon, G.
The sequence of Homo sapiens BAC clone GSI-405L21
Unpublished
2 (bases 1 to 235141)
Waterston, R.
Direct Submission
Submitted (24-APR-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
3 (bases 1 to 235141)
Waterston, R.
Direct Submission
Submitted (03-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 1
AUTHORS
JOURNAL
TITLE
REFERENCE
AUTHORS
JOURNAL
TITLE
COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.edu
Summary Statistics
Center project name: H_GS405L21

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems,

such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. Michael Lovett, Departments of Otorhinolaryngology, Molecular Biology and Oncology, University of Texas Southwestern Medical Center, Dallas TX. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).
 Cell line: Lymphoblastoid

Haplotypes: two
 VECTOR: pBelobAC
 Selection: chloramphenicol
 NEIGHBORING SEQUENCE INFORMATION:
 The actual start of this clone is at base position 1 of GSI-405L21; actual end is at 235141 of GSI-405L21.

FEATURES
 This clone contains STS HSA0552D9 (MID:g1233025).
 Location/Qualifiers

```

source 1..235141
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5p15.2"
/clone="GSI-405L21"
/clone_1lb="GSBAC1"
181..480
/rpt_family="Alu"
3000..3303
/rpt_family="Alu"
4356..4559
/rpt_family="TGA)n"
4576..4744
/rpt_family="MER1_type"
6769..7074
/rpt_family="Alu"
7096..7157
/rpt_family="Alu"
7286..7384
/rpt_family="(CA)n"
7491..7742
/rpt_family="MaLR"
7915..8234
/rpt_family="MaLR"
9154..9234
/gene="SEMAF"
join(9154..9277,51204..51303,70597..70642,151152..151214,162076..162174,164155..164368,186819..187104,191786..191921,196574..198778,234207..234415)
/note="match to U52840 (PID:22772584); but note polymorphism at position 56; H_GS405L21.1"
/codon_start=1
/product="Semaphorin F"
/protein_id="AAC1468.1"
/db_xref="GI:3080661"
/translation="MKGTCVIAIMFSSGLIMRLAHPDAOGTTQCO RTEHPVTSYEIG PWLREFRAKNVADFSOLFDPGOKELVGARVYLPFLOEDLSLQAVDEDEATRK ACYSKSKSEBCQNTATRVLLVGGDLFCGCTAFTPCVTCNRSLSNTEIHDISGAR CPYSPOHNSTALTITAGGELIATATPANDFPERDPAIYRSGLILPPLTAQYNSRWLARP NVSSYDIGNFTYFFRENNAVERDCKTVESRAKVCNDIGRFLIEDTWFPMKARL NCSRGEVFFYVNEIQLSTFELDLIVGIFTNNNSIAASAVCFNLSAIAQASGP FYOENSRGAMLPYNPDPHFOCGVLDGLVNLERNLQDOKFLLIMEVQVPTTV PSEMDNSRSHVADVVGREAIYHIITLATDYGTKIKRVRLPMTQSSSCLLEIEL PEERRRPIRSQIHSQSVLTVGLREHVTKIPLAKRCQYRTRR"
10974..11242
/rpt_family="L1"
11271..11361
/rpt_family="L1"
12176..12392
repeat_region
repeat_region
repeat_region

```

```

repeat_region      /rpt_family="L1"
                    13349..13391
repeat_region      /rpt_family="MER81"
                    13453..13898
repeat_region      /rpt_family="MaLR"
                    19001..19202
repeat_region      /rpt_family="L2"
                    19350..19841
repeat_region      /rpt_family="L1"
                    19893..20700
repeat_region      /rpt_family="L1"
                    21002..21401
repeat_region      /rpt_family="MaLR"
                    21797..21994
repeat_region      /rpt_family="MER1_type"
                    22413..22710
repeat_region      /rpt_family="Alu"
                    27381..27411
repeat_region      /rpt_family="Mariner"
                    27785..28082
repeat_region      /rpt_family="Alu"
                    29733..29893
repeat_region      /rpt_family="MIR"
                    30444..30532
repeat_region      /rpt_family="L2"
                    33326..33653
repeat_region      /rpt_family="L2"
                    33961..34048
repeat_region      /rpt_family="MER1_type"
                    34057..34316
repeat_region      /rpt_family="L2"
                    34390..34520
repeat_region      /rpt_family="L2"
                    36072..36286
repeat_region      /rpt_family="MER1_type"
                    36464..36604
repeat_region      /rpt_family="Alu"
                    36622..36732
repeat_region      /rpt_family="(TAGA)n"
                    38061..38641
repeat_region      /rpt_family="MER21_g"
                    38939..39082
repeat_region      /rpt_family="MER1_type"
                    39096..39399
repeat_region      /rpt_family="Alu"
                    39415..39483
repeat_region      /rpt_family="MER1_type"
                    39557..39683
repeat_region      /rpt_family="L2"
                    42462..42767
repeat_region      /rpt_family="Retroviral"
                    42862..43091
repeat_region      /rpt_family="Retroviral"
                    46106..46225
repeat_region      /rpt_family="MIR"
                    47785..47839
repeat_region      /rpt_family="(CA)n"
                    51495..51731
repeat_region      /rpt_family="MIR"
                    53061..53183
repeat_region      /rpt_family="L2"
                    54138..54238
repeat_region      /rpt_family="MER53"
                    54275..54314
repeat_region      /rpt_family="L2"
                    54348..54495
repeat_region      /rpt_family="MIR"
                    58611..58868
repeat_region      /rpt_family="Alu"
                    59061..59261
repeat_region      /rpt_family="MER2_type"
                    59469..59532
repeat_region      /rpt_family="L2"

```

```

repeat_region      59571..59768
repeat_region      /rpt_family="MER1_type"
                    60224..60521
repeat_region      /rpt_family="Alu"
                    62226..62479
repeat_region      /rpt_family="Alu"
                    62480..62757
repeat_region      /rpt_family="Alu"
                    63954..64048
repeat_region      /rpt_family="MIR"
                    64650..64950
repeat_region      /rpt_family="Alu"
                    65069..65357
repeat_region      /rpt_family="Alu"

```

```

Query Match          2.0%; Score 20; DB 85; Length 235141;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 271 cccagcccccagagagaaa 290
Db 156934 cccagcccccagagagaaa 156915

```

```

RESULT 39
AF178954/c
LOCUS AF178954 2051 bp mRNA 19-OCT-2000
DEFINITION Mus musculus monocyte diacylglycerol transporter 4 mRNA, complete cds.
ACCESSION AF178954
VERSION AF178954.1 GI:10880481
KEYWORDS
SOURCE
ORGANISM
Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS Yoon,H. and Philip,N.J.
TITLE 1 (bases 1 to 2051)
JOURNAL Cloning and expression of mouse MCT3 and MCT4
REFERENCE
AUTHORS Yoon,H. and Philip,N.J.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1999) Cell and Molecular Biology, Pennsylvania
College of Optometry, 8360 Old York Road, Philadelphia, PA 19027,
USA

```

```

FEATURES
source
Location/Qualifiers
1..2051
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/tissue_type="skeletal muscle"
1..1413
/note="MCT4"
/codon_start=1
/product="monocarboxylate transporter 4"
/protein_id="AA024271.1"
/db_xref="GI:10880482"
/translation="MGCAVVDGPGPIKAPDGGGMVAFPGCFITITGSPVAPKAVSV

```

```

CDS
1..1413
/note="MCT4"
/codon_start=1
/product="monocarboxylate transporter 4"
/protein_id="AA024271.1"
/db_xref="GI:10880482"
/translation="MGCAVVDGPGPIKAPDGGGMVAFPGCFITITGSPVAPKAVSV

```

```

BASE COUNT 372 a 602 c 608 g 469 t
ORIGIN

```

```

Query Match          1.9%; Score 19; DB 94; Length 2051;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 264 agagacgagccagccagc 282
|||||
Db 74 AGGACGGCCAGCCAGC 56

RESULT 40
LOCUS RN087627/c
DEFINITION Rattus norvegicus putative monocarboxylate transporter (MCT3) mRNA,
complete cds.
ACCESSION U87627
VERSION U87627.1 GI:2463650
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2118)
AUTHORS Wilson, M.C., Jackson, V.N., Heddle, C., Price, N.T., Pilgaard, H.,
Juel, C., Bonen, A., Montgomery, I., Hutter, O.F. and Halestrap, A.P.
TITLE Lactic acid efflux from white skeletal muscle is catalyzed by the
monocarboxylate transporter isoform MCT3
J. Biol. Chem. 273 (26), 15920-15926 (1998)
98298091
2 (bases 1 to 2118)
AUTHORS Jackson, V.N., Price, N.T. and Halestrap, A.P.
TITLE Direct Submission
Submitted (29-JAN-1997) Cellular Biochemistry, Hannah Research
Institute, Mauchlin Road, Ayr KA6 5HL, UK
FEATURES
source location/Qualifiers
1..2118
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
/dev_stage="12 weeks"
/sex="male"
/tissue_type="skeletal muscle"
1..2118
/gene="MCT3"
90..1505
/note="similar to Gallus gallus retinal epithelial
membrane protein, encoded by GenBank Accession Number
U13685; similar to Human MCT3, encoded by GenBank
Accession Number U81800"
/product="putative monocarboxylate transporter"
/protein_id="AAC53591.1"
/db_xref="GI:2463651"
/translation="MGAAVVDGPGPIGAKPDGGMGMAVLEGGFIITGSGYAPKAVSV
FEKLMHEFGIYSDTAMISILLAMLYGTGLSCMVCNRFRCRPMLVGLFASIGM
VAASFCRSIIQIYLTGTYTGLGLNFPGLIMLNRYNRRPMAANGAAGSPVFL
CALSPGLOLDHGMRGGLILGLGLNLCVCAALMRPLVAPASGAEPPGPS
PRLILSVFRORGLIYVAASIMVGLFVPPVYVYAKDMGVPDTRKAPLITLIGF
IDIFARPTAGFTTGKAKRPRYSYVLFSPAFPNNGTDLGTSASDYGCLVVCIRFGI
STVGALQDFEVLMAIVGOKFSSAIGVLLEAAVAVLIGPSGKLDLATVYRVE
ILAGAEVLSVLILLGNFCLGKRRREPVTPEVASEEKLHKKPPVDVVDVRSVE
HFLKAPKENGVTPTPTSTV"

BASE COUNT 401 a 605 c 629 g 483 t
ORIGIN

Query Match 1.9%; Score 19; DB 95; Length 2118;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 264 agagacgagccagccagc 282
|||||
Db 163 AGGACGGCCAGCCAGC 145

RESULT 41
LOCUS AF136600/c
DEFINITION Magnaporthe grisea protein kinase C (mpkc) gene, complete cds.
ACCESSION AF136600
VERSION AF136600.1 GI:4928704
KEYWORDS
SOURCE Magnaporthe grisea.
ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes Incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 5841)
AUTHORS Khang, C.H. and Lee, Y.H.
TITLE The identification and characterization of protein kinase C, mpkc,
in Magnaporthe grisea
Unpublished
2 (bases 1 to 5841)
AUTHORS Khang, C.H. and Lee, Y.H.
TITLE Direct Submission
Submitted (23-MAR-1999) Division of Applied Biology and Chemistry,
Seoul National University, 103, Seodun-dong, Kwonson-gu, Suwon 441
-744, South Korea
FEATURES
source location/Qualifiers
1..5841
/organism="Magnaporthe grisea"
/strain="70-15"
/db_xref="taxon:148305"
/note="Submitted through Biological Research Information
Center of Korea GeneNuri No. KS101751"
join(<883..1342,1427..1693,1772..3923,3987..4188,
4242..4647,4716..>4777)
/gene="mpkc"
/product="protein kinase C"
/codon_start=1
/product="protein kinase C"
/protein_id="AAD33693.1"
/db_xref="GI:4928705"
/translation="MDDRLODIYKKIEREKALINANLMROOTNDNAVSKLDTORE
GRNLEFFEGTILREMOGRMGOGMDNLSTGCTLAASASSSSRPSAEDEDPMPPK
DCGYAGDGDTOTYSQIGEGDLMPPGPPANGPGRSHIKTRPNTFKLLIKDT
PYLGPRIQLMISQIQFKNVEEQYLKIEKMYQLQMEGDKSRDAARVRESOKI
VLLKQALRYEELHIDISGDSDDSIIMPMILRKLPLGQLSIRVLAIDVDHAPGR
FARGPDTFVAIKVEDNIARTVRSRTDRENEYHNVYDVKANEIELTYDKGEHSIP
IGMLWYRIIDIAEELRRKIEAETNSLGVNSADRMGDVPAARVPPPPMNTGROP
GGPVSPGLOGOOGGYGDPDPOOGPOGGOVOPITGMFLLELTGSLTSLNFKO
SKDRRPVDLGLGRKATIRKKEVEHMGHKTVEROFYINRCALCGDLKSGSMOC
EDCKYTCIKCYTSVATKTSKSNMETDDEERINRIPHRVAFNSNLNANNCCHGY
LILPFRKNCRKSECOMAHACVHLVDFPCMSMAVANQIIEGIRSORQOKASM
SDRTLRSGKSPSGSHASSAFSGOMSGSYGOASPEATEAKFMYSQTSQRTSP
DRTSSSSQAAATAMTGLPGAMSOPROGCOOPTTVDGASGGYSGYGPDDPYA
GPOOSPPPOQAAYGPOORRYKPNADYANISGCGSPQMAOQPOARPOQOQOPLY
SPQOASQAQOQPLSPYKQDHOEQIISPAAGVYIPRSARPLPSATDETCGRIGDH
ENFLVLVGLNGKWLMAESKTRKLYALVKLKEETIENDVESIRSKRVFLANK
ENHPLETLNHACFOETRYVEEYISGDLMLHLIQRGFGTRAFQVFAEVLAKY
PHENGVIYEDLKLNDLILTLDDGHIKADYGLCKEDMYSTSTGCGYEPMAPELIL
DKYGAAYWMAFGLIYOMLLOOSPFREDEDEYDALADEPLVPIIMPDSVSL
OKLTIREPDRIGSGPTQAOEYMSOEFENIYWDIDYHRAVAPFLPQIKSNTDINSF
DSEFTSVTVLIPVOSVLSQAMOEFRGKSTYADPD"

BASE COUNT 1447 a 1612 c 1479 g 1303 t
ORIGIN

Query Match 1.9%; Score 19; DB 13; Length 5841;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 413 gaaggacatttgagcagc 431
|||||

Db 464 GAGGACATTGGCAGC 446

RESULT 42

CEL32B5

LOCUS CEL32B5 42545 bp DNA INV 05-DEC-1996

DEFINITION Caenorhabditis elegans cosmid C32B5.

ACCESSION U80843

VERSION U80843.1 GI:1707218

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berts, M., Bonfield, J., Burton, J., Connell, M., Copesey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A., Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. and Wohlman, P.

TITLE

2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

JOURNAL

MEDLINE

94150718

REFERENCE

2 (bases 1 to 42545)

AUTHORS

Scheet, P.

TITLE

The sequence of C. elegans cosmid C32B5

JOURNAL

REFERENCE

3 (bases 1 to 42545)

AUTHORS

Waterston, R.

TITLE

Direct Submissions

JOURNAL

COMMENT

Submitted by:

Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: r.wenematode.wustl.edu and jesesanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is R07C3. 200 bp overlap; 3' cosmid is F45D11. Actual start of this cosmid is at base position 1 of CEL32B5; actual end is at 42545 of CEL32B5

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

Location/Qualifiers

1. .42545

/organism="Caenorhabditis elegans"

gene

CDS

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="II"

/clone="C32B5"

complement(538..2159)

/gene="C32B5.10"

complement(join(538..670,762..1790,2089..2159))

/gene="C32B5.10"

/codon_start=1

/evidence=not_experimental

/protein_id="AAB37966.1"

/db_xref="GI:1707228"

/translation="MAYSLPTPLPSLCKVIKHMENVSRIOISQPCPSIRKAENSIPLKIGHELQOCRKVINNTDYSLAVVCNFPSSPYDLDEVSMDLGEVIDQRDVPVNDSEL EGEGRILRMREYELRMLDPMCKKALIGELKLEKMTTEKSELVNRKCVGIKISLFLSRVGVPAPIMDIDFVNKNKLTTPEDDKSYMKLHOAPKYLIFKLFSGRRADIOVKLSIEKNGILRVASLNLKVTHELIEIGSGYDKAISFOPLBESSLSQISKINDPMIENPLQTSPLTIFKNPNENLFRASATFTNVRVELEKSTIFPREALDIAMVLRKGEVGTINWFSLTEVTALVCFKDIRKLPGARYAGHKTSQDGRSVLIPINAFSEIHIDEKEEAQGNMWSLKITVRRKL"

complement(3176..3625)

/gene="C32B5.11"

complement(join(3176..3446,3500..3545,3589..3625))

/gene="C32B5.11"

/codon_start=1

/evidence=not_experimental

/protein_id="AAB37965.1"

/db_xref="GI:1707227"

/translation="MYIEMLGVSVHFVNNQDPAFCYSRRHNRNLELLDOKAESFWSKFRVLDPKRSIPFKSLDLNGSTIDDNVIOSELLIHNGRIDENIDLEFKLIANQOVELITTVYGGSPMMVEL"

complement(4928..8090)

/gene="C32B5.12"

complement(join(4928..5078,5462..6453,7974..8090))

/gene="C32B5.12"

/codon_start=1

/evidence=not_experimental

/protein_id="AAB37967.1"

/db_xref="GI:1707229"

/translation="MWSLSTLTITLPPKNFLSKYFIASWSPLSFKSSCPGLAIPVAKCPSLRAEKSVPLRIKHIRTPTPTQINTROLAIGCGDNPSPSYDIDQGEODLGKLRPASDDEITLIDGRRHETEEERKANRFEVAAIIRENRLAFAPADAVETVEPT EVRRPNHGGOLVNRNSQLESVIPSAMWHTFTQARGLTEPRLSKGGMGRKLOEKLIEVRVATKRLHEIMKYLAEKLEFGKHLRIQKLEIGSNPGVRLPNMMTIFRNLELPE DHEVLKSIIRPLMLKSLPLASLHPLPLESIVITPSKLLSELIDVDFEPEIKKKNOEVYAKPEDEARAHFESPEVYAGSSEARQATEPTPTFNIMKSSSXAENVS IHCELPNLRHNVFSMTVOKPELIID"

8026..10315

/gene="C32B5.9"

join(8026..8540,9362..10138,10195..10315)

/gene="C32B5.9"

/codon_start=1

/evidence=not_experimental

/protein_id="AAB37964.1"

/db_xref="GI:1707226"

/translation="MYILKKFEGGRVIVKDKLTIFCRILVRIPASLKQVHNLQFSDNNQLPEAKPLIDFSTPLSITLFGKVIINVEDPIRSTETLYDGFLLIDNL ENLKLKRRKRLHSLSKSSRQNPRLIDMWIEGRAGPEVSYSEMMEEIRARPIIKTVMINKERITIERHLSQOYPSLSTERSVPLRLDSLKLGNSDVSNTICELAYIRHGGRLSKKRKLIDVTEGKVDNMIVVEGELIIVLDRGREINTADVAKIFEGRGILITVGTLSIGCRGILIRPSRLKFEIRDLDEGSEDNKLEIEIKRPLHLTSPFNFSFNSTNLRVEDPYOOSTGVLIIFKNISSEFNNELENELHKKRAHLSFDREYFELQIIVIGLGM IKFGRDIGITYSVELQAKNGSEVLEIIRKHHERIIDRPEDPNIVLLRMNNLALYISYEFYKCYCMHRVQPR"

10652..11939

/gene="C32B5.8"

join(10652..10716,11060..11572,11628..11756,11807..11939)

/gene="C32B5.8"

/codon_start=1

/evidence=not_experimental

/protein_id="AAB37957.1"

/db_xref="GI:1707219"

/translation="MEKRPKLSYASTVIOYSDPMLRRLSQQCSLSIGSTDKSVPLHLE

```
SLKLSGNSISVNALEYELAIYRHGARGPKKFKVDYTERKGRVDMNLVEDEPEILLVDL
RGNORITANVYTELANORRIENIEKMRNNYPLTVLVGVKRMVEMIEYERKLHFAVK
YLVEKLLGGRIIGKIVDLISRSRGLIRIIPSRILKRSLLPGRVHMMHAGREIGSYATFVI
TEQNIENEMASRIREFEIGIANKTGETITNMKADSOVLITKIQGDAMILGMKVKSRV
LEI"
gene
/12939..14079
/gene="C32B5.16"
join(12939..13009,13064..13903,13953..14079)
/gene="C32B5.16"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB37971.1"
/db_xref="GI:1707233"
/translation="MSYKPMSPYPIKCLLEHLPNLRPKLALSGPLIKNLEKIYHL
IKYLIKIDHSTINKDPTQYPTVHNKNGFLIDHIDRDLDSPIHEDNEELIIVKRN
IRWIFRANVSSSENTIQIKDKFEITLRRKMHALKYLGLKEGTVKYNHLITIK
CGNVIIRILSTVKEFKVILVEIDENTYDELDTIRSLIQTSPLANEITRFLOAPSRH
ELIRTAGMLERICSIDVHTLATLNHRVHABYFYEMRNSCFVDFLEIMMEYGEKIGSY
YTFNIFEHRIIDKAMKRIKMYKRSVEAGSARRAGQLTLNMNNSQLVTSKYVGDAWI
LGVEYKDRMVL"
complement(15330..15886)
gene
/15330..15886
/gene="C32B5.13"
complement(join(15330..15493,15542..15723,15780..15886))
/gene="C32B5.13"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB37968.1"
/db_xref="GI:1707230"
/translation="MYAKANNRYLVSEEQIIDCGNFTSPQENILSHEPIKNGV
TEADYPYKGENEKCKYDENKIKLMPNTMLVGNLPETILKFIEMHGPYFRMKAP
SEFNYKIGVSPTEOECATDARSILTVGIEGGQMYWIKSGSFG"
15984..18023
/gene="C32B5.7"
join(15984..17051,17217..17561,17627..17802,17860..18023)
/gene="C32B5.7"
/note="strong similarity to the peptidase family C1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB37963.1"
/db_xref="GI:1707225"
/translation="MSKPSYSQSLKTVLKHVDENLRKFKPKPANLBNTRYIGLFL
LTALEFYIGALVOORTQNYKNKPKFLDMWDEGVGVKDGQNCNASYAFALISAIE
SMVAIANQQLSFSEQIIDCGAIESDPMAMWTYERKGIETTYTFEYKGNK
CEYDSKRAYLILDDTYDMSDESLALVFIDERPGPLFTNTPSPSFENYSGIYNPEEE
CKSTNEKALITVCGNDKGONWYIVKSGSFGT"
complement(19648..20937)
gene
/19648..20937
/gene="C32B5.14"
complement(join(19648..19762,19822..20814,20867..20937))
/gene="C32B5.14"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB37969.1"
/db_xref="GI:1707231"
/translation="MSNSKPVNYTLKCLKNMDSGMRLQLFVMPSTIOYLEKLEPLH
VKYLTIKSDHITVNTTFOPLKCNKYNMFHGGYKDFPOYGLDGEIDDESDIID
VRDGLSKKGIJCEMEETARLVNLTLOKSORYSRWISHSHTILKLSIYVPEENFI
RLTKGKREVEYOKRIEAMKYLILGPGRSLEANPSTIGCDYLVLPSTLKRRIE
NLITPSEKIANTLDVNOIVDNSSPLSLKTSFENHLYHHPHSLVKTVKMKLLEVM
VPDIYSGIVSNLQWVDEKRAHIVFLGDCSTSNFLKILAHMLFEHRDITGYHTYOLSE
```

```
Query Match 1.9%; Score 19; DB 6; Length 42545;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 gctcccaactccagcag 235
|||||
Db 5471 GCTTCCCACTTCCAGCAG 5489

RESULT 43
AC007191 43390 bp DNA PRI 02-APR-1999
LOCUS AC007191
DEFINITION Homo sapiens chromosome 19, cosmid R34072, complete sequence.
```

```
AC007191
VERSION AC007191.1 GI:4558643
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 43390)
Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Viswanathan,V.,
Burkhardt-Schultz,K.U., Gordon,L., Dias,D., Ramirez,M.,
Stillwagen,S., Phan,H., Velasco,N., Do,L., Regala,W., Terry,A.,
Garnes,J., Dangnanan,L., Erler,A., Christensen,M., Georgescu,A.,
Avila,J., Liu,S., Attix,C., Andreise,T., Trankheim,M.,
Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R.,
Thomas,P., Quan,G., Krommiller,B., Ariellano,A., Sanders,C., Ow,D.,
Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carrino,A.V.
Sequence analysis of a 1.9 Mb region in 19q13.2 between APOE and
D19S412
Unpublished
2 (bases 1 to 43390)
Lamerdin,J.E.
Direct Submission
Submitted (02-APR-1999) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
Map and sequence oriented from q centromere to telomere. Cosmid
R34072 overlaps cosmid R28204 (AC006132) to the left from bases 1
to 2,155 of this accession, and overlaps cosmid R26955 to the right
from bases 37,677 to 43,390. Additional chromosome 19 map and
sequence information may be obtained at:
ftp://www-bio.1lnl.gov/bdrp/genome/genome.html.
Location/Qualifiers
1..43390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R34072"
/chromosome="19"
/map="19q13.2 between APOE and D19S412"
/cell_line="5HL2-B"
/clone_lib="L19NC03 R chromosome 19-specific cosmid
library"
/note="Cosmid library constructed at LINT from flow-sorted
chromosomes from human-hamster hybrid 5HL2-B, which
carries chromosome 19 as its only human chromosome."
35..306
/rpt_family="AluSg"
/rpt_family="AluSg"
307..474
/rpt_family="AluJc/FRAM"
477..787
/rpt_family="AluSx"
complement(835..976)
/rpt_family="MIR"
1142..1443
/rpt_family="AluSg"
join(1533..1574,1658..2359)
/gene="GIPR"
/product="Human GIP receptor (GIPR) mRNA, partial cds."
4533..2359
/gene="GIPR"
/note="gastric inhibitory polypeptide receptor"
join(1533..1574,1658..1864)
/note="GASTRIC INHIBITORY POLYPEPTIDE RECEPTOR PRECURSOR;
GIP-R; GLUCOSE-DEPENDENT INSULINOTROPIC POLYPEPTIDE
RECEPTOR"
/codon_start=1
/product="GIPR_HUMAN [AA 385- 466]"
/protein_id="AAD22674.1"
/db_xref="GI:4558645"
/translation="GFIVSVLYCFINKEVSEIRGWHCHRLRSLSGEQQLPRPAR
RALPSSGSGPEVPTSRGLSSGTLPEGVNASEKLELSYC"
complement(1590..11709)
/rpt_family="CGG)n"
complement(1886..2236)
```

```

repeat_region      /rpt_family="LINE2"
                   complement(2357..2492)
repeat_region      /rpt_family="Aluub"
                   complement(2493..2771)
repeat_region      /rpt_family="Alusx"
                   complement(2786..3192)
misc_feature       /rpt_family="LINE2"
                   3223..3544
                   /note="Dds similarity to overlapping ESTs:
                   (3223..3435) AA345686 EST51736 Gall bladder I Homo sapiens
                   cDNA 5' end; Score: 411 Identity: 211/214 (98%).
                   (3544..3253) AA987372 or92h03.s1 NCI_CGAP_Lu5 Homo sapiens
                   cDNA clone IMAGE:1603349 3'; Score: 575 Identity: 291/291
                   (100%).
                   (3544..3266) AA83616 ne40e08.s1 NCI_CGAP_C03 Homo sapiens
                   cDNA clone IMAGE:899846; Score: 550 Identity: 277/279
                   (99%)."
repeat_region      4318..4431
                   /rpt_family="MIR"
repeat_region      4440..4738
                   /rpt_family="Alusg"
                   complement(4916..5215)
repeat_region      /rpt_family="AlusC"
                   5362..5603
repeat_region      /rpt_family="Aluub"
                   5766..5995
repeat_region      /rpt_family="MIR"
                   complement(6052..6349)
repeat_region      /rpt_family="Alusx"
                   complement(6428..6469)
repeat_region      /rpt_family="Alu"
                   complement(6656..6942)
repeat_region      /rpt_family="Alusx"
                   complement(6943..7077)
misc_feature       /rpt_family="Alusg/x"
                   complement(7381..7656)
                   /note="Dds similarity to overlapping ESTs:
                   (7656..7381) AA983186 oq52h06.s1 NCI_CGAP_Kids Homo
                   sapiens cDNA clone IMAGE:1590011 3' similar to
                   SW:SMD2_HUMAN P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM
                   D2; (270..1) 99% identity.
                   (7656..7388) AA644001 nm19e12.s1 NCI_CGAP_C010 Homo
                   sapiens cDNA clone IMAGE:1060654 similar to SW:SMD2_HUMAN
                   P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2; (269..1)
                   99% identity.
                   (7656..7382) A1092009 oz97a08.x1
                   Soares.parathyroid.tumor_NBHPA Homo sapiens cDNA clone
                   IMAGE:1683254 3' similar to SW:SMD2_HUMAN P43330 SMALL
                   NUCLEAR RIBONUCLEOPROTEIN SM D2; (275..1) 98% identity.
                   (7656..7391) W58351 zd19h08.r1 Soares fetal heart NBHH19W
                   Homo sapiens cDNA clone 341151 5' similar to SW:SMD2_HUMAN
                   P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2; (212..478)
                   99% identity.
                   and others."
repeat_region      complement(join(7381..7656,8316..8495,11816..11844))
                   /product="Human SnRNP core protein Sm D2 mRNA, complete
                   cds"
                   complement(join(7482..7656,8316..8495,11816..11817))
                   /function="mRNA SPLICING"
                   /note="SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2; SNRNP CORE
                   PROTEIN D2; SM-D2"
                   /product="SMD2_HUMAN"
                   /codon_start=1
                   /protein_id="AAD22673.1"
                   /db_xref="GI:4558644"
                   /translation="MSLINKPKSEMTPELOKREEEFNTPSLVLTQSVKNNTOVLI
                   NCRNNKRLGKVAKAFDRRCNNVLENVKEMTVEPKSGKRRKSPVKNDRYISKMFLR
                   GDSVIVLRNPLIAGK"

```

```

repeat_region      7803..7912
                   /rpt_family="MIR"
repeat_region      complement(7923..8211)
                   /rpt_family="Alusx"
misc_feature       complement(8316..8497)
                   /note="Dds similarity to overlapping ESTs:
                   (8497..8316) AA983186 oq52h06.s1 NCI_CGAP_Kids Homo
                   sapiens cDNA clone IMAGE:1590011 3' similar to
                   SW:SMD2_HUMAN P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM
                   D2; (457..276) 99% identity.
                   (8491..8316) A1004535 ot63e01.s1 Soares testis_NHT Homo
                   sapiens cDNA clone IMAGE:1621464 3' similar to
                   SW:SMD2_HUMAN P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM
                   D2; (446..271) 99% identity.
                   (8497..8316) AA644001 nm19e12.s1 NCI_CGAP_C010 Homo
                   sapiens cDNA clone IMAGE:1060654 similar to SW:SMD2_HUMAN
                   P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2; (451..270)
                   99% identity.
                   (8495..8316) A1092009 oz97a08.x1
                   Soares.parathyroid.tumor_NBHPA Homo sapiens cDNA clone
                   IMAGE:1683254 3' similar to SW:SMD2_HUMAN P43330 SMALL
                   NUCLEAR RIBONUCLEOPROTEIN SM D2; (455..276) 100%
                   identity.
                   (8495..8316) W58351 zd19h08.r1 Soares fetal heart NBHH19W
                   Homo sapiens cDNA clone 341151 5' similar to SW:SMD2_HUMAN
                   P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2; (33..211)
                   98% identity.
                   and others."
repeat_region      complement(8666..8961)
                   /rpt_family="AlusC"
repeat_region      8962..9315
                   /rpt_family="LINE2"
repeat_region      9319..9446
                   /rpt_family="Alusx"
repeat_region      9447..9477
                   /rpt_family="Alusx"
repeat_region      9487..9661
                   /rpt_family="CAAAA)n"
repeat_region      9681..9975
                   /rpt_family="Alusx"
repeat_region      9979..10278
                   /rpt_family="Alusx"
repeat_region      10279..10433
                   /rpt_family="Alu"
repeat_region      10479..10780
                   /rpt_family="Alusg/x"
repeat_region      10820..10903
                   /rpt_family="LINE2"
repeat_region      complement(11059..11360)
                   /rpt_family="Alusx"
repeat_region      complement(11467..11599)
                   /rpt_family="FLAM_C"

Query Match      1.9%; Score 19; DB 86; Length 43390;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 ccacagattcaaacccag 25
Db      29185 CCACAAATCAACCCAG 29203

RESULT 44
AP000333      47404 bp      DNA      PRI      20-NOV-1999
LOCUS      Homo sapiens genomic DNA, Chromosome 21q22.1, D21S226-AML region,
DEFINITION      clone:B333B2, complete sequence.
ACCESSION      AP000333
VERSION      AP000333.1 GI:4835702
KEYWORDS      HNC. Homo sapiens DNA, clone:B333B2.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE 1 (bases 1 to 47404)
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., and Sakaki,Y.
 TITLE Homo sapiens 47,404bp genomic DNA of 21q22.1
 JOURNAL Published Only in Database (1999) In press
 REFERENCE 2 (bases 1 to 47404)
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., and Sakaki,Y.
 JOURNAL Submitted (13-MAY-1999) to the DDBJ/EMBL/Genbank databases.
 TITLE Masahito Hattori, The Institute of Physical and Chemical Research
 (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail:hattori@gscc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
 Tel:81-42-778-9923, Fax:81-42-778-9924)
 COMMENT The sequence is a part of the data (ACCESSION No. AP000165 -
 AP000173).
 The sequencing project is supported by Japan Science Technology
 Corporation (JST) and The Institute of Physical and Chemical
 Research (RIKEN).
 FEATURES
 source Location/Qualifiers
 1..47404
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /clone="B353B2"
 /map="21q22.1"
 /map="21q22.1"
 BASE COUNT 13800 a 9563 c 10009 g 14032 t
 ORIGIN
 Query Match 1.9%; Score 19; DB 91; Length 47404;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 641 ccagtcacgaagagctcgt 659
 ||||||||||||||||
 Db 8162 CCAGTCCAGGAGCTCTG 8180
 RESULT 45
 AC091020/c
 LOCUS Homo sapiens chromosome 18 clone RP11-701C7 map 18, LOW-PASS
 DEFINITION
 AC091020 61435 bp DNA HTG 24-MAR-2001
 SEQUENCE SAMPLING.
 AC091020
 AC091020.1 GI:13443176
 KEYWORDS HTG: HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 61435)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., and Lander,E.
 TITLE Homo sapiens chromosome 18, clone RP11-701C7
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 61435)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B., Brown,A.,
 Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
 Collins,S., Collymore,A., Cooke,P., Dearrellano,K., Dewar,K.,
 Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,M.,
 Iliev,I., Johnson,R., Jones,C., Karatas,A., Larocque,K.,
 Lamazares,R., Landers,T., Lehotsky,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Margulis,N., Matthews,C., McCarthy,M.,
 McEwan,P., McKernan,K., McPheters,R., Meldrum,J., Menus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,

REFERENCE 1 (bases 1 to 47404)
 AUTHORS Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
 Severy,P., Sougnier,C., Spencer,B., Stenge-Thoman,N.,
 Sciojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S.,
 Theodore,J., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A., and Zody,M.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 COMMENT
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L12249
 Center clone name: 701_C7

 * NOTE: This record contains 75 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 1
 683 782: contig of 682 bp in length
 783 782: gap of 100 bp
 1496 1595: contig of 713 bp in length
 1596 2335: gap of 100 bp
 2336 2435: contig of 740 bp in length
 2436 3147: gap of 100 bp
 3148 3247: contig of 712 bp in length
 3248 3964: gap of 100 bp
 3965 4064: contig of 717 bp in length
 4065 4755: gap of 100 bp
 4756 4855: contig of 691 bp in length
 4856 5569: gap of 100 bp
 5570 5669: contig of 714 bp in length
 5670 6349: gap of 100 bp
 6350 6449: contig of 680 bp in length
 6450 7167: gap of 100 bp
 7168 7267: contig of 718 bp in length
 7268 7989: gap of 100 bp
 7990 8089: contig of 722 bp in length
 8090 8828: gap of 100 bp
 8829 8928: contig of 739 bp in length
 8929 9638: gap of 100 bp
 9639 9738: contig of 710 bp in length
 9739 10460: gap of 100 bp
 10461 10560: contig of 722 bp in length
 10561 11296: gap of 100 bp
 11297 11396: contig of 736 bp in length
 11397 12093: gap of 100 bp
 12094 12193: contig of 697 bp in length
 12194 12905: gap of 100 bp
 12906 13005: contig of 712 bp in length
 13006 13725: gap of 100 bp
 13726 13825: contig of 720 bp in length
 13826 14540: gap of 100 bp
 14541 14640: contig of 715 bp in length
 14641 15287: gap of 100 bp
 14641 15287: contig of 647 bp in length
 15288 15387: gap of 100 bp
 15388 16101: contig of 714 bp in length
 16102 16201: gap of 100 bp
 16202 16876: contig of 675 bp in length

```
* 16877 16976: gap of 100 bp
* 16977 17740: contig of 764 bp in length
* 17741 17840: gap of 100 bp
* 17841 18522: contig of 682 bp in length
* 18523 18622: gap of 100 bp
* 18623 19353: contig of 731 bp in length
* 19354 19453: gap of 100 bp
* 19454 20186: contig of 733 bp in length
* 20187 20286: gap of 100 bp
* 20287 21003: contig of 717 bp in length
* 21004 21103: gap of 100 bp
* 21104 21821: contig of 718 bp in length
* 21822 21921: gap of 100 bp
* 21922 22634: contig of 713 bp in length
* 22635 22734: gap of 100 bp
* 22735 23430: contig of 696 bp in length
* 23431 23530: gap of 100 bp
* 23531 24226: contig of 696 bp in length
* 24227 24326: gap of 100 bp
* 24327 25057: contig of 731 bp in length
* 25058 25157: gap of 100 bp
* 25158 25918: contig of 761 bp in length
* 25919 26018: gap of 100 bp
* 26019 26753: contig of 735 bp in length
* 26754 26853: gap of 100 bp
* 26854 27578: contig of 725 bp in length
* 27579 27678: gap of 100 bp
* 27679 28378: contig of 700 bp in length
* 28379 28478: gap of 100 bp
* 28479 29202: contig of 724 bp in length
* 29203 29302: gap of 100 bp
* 29303 30065: contig of 763 bp in length
* 30066 30165: gap of 100 bp
* 30166 30876: contig of 711 bp in length
* 30877 30976: gap of 100 bp
* 30977 31677: contig of 701 bp in length
* 31678 31777: gap of 100 bp
* 31778 32495: contig of 718 bp in length
* 32496 32595: gap of 100 bp
* 32596 33291: contig of 696 bp in length
* 33292 33391: gap of 100 bp
* 33392 34102: contig of 711 bp in length
* 34103 34202: gap of 100 bp
* 34203 34972: contig of 770 bp in length
* 34973 35072: gap of 100 bp
* 35073 35795: contig of 723 bp in length
* 35796 35895: gap of 100 bp
* 35896 36671: contig of 776 bp in length
* 36672 36771: gap of 100 bp
* 36772 37483: contig of 712 bp in length
* 37484 37583: gap of 100 bp
* 37584 38290: contig of 707 bp in length
* 38291 38390: gap of 100 bp
* 38391 39091: contig of 701 bp in length
* 39092 39191: gap of 100 bp
* 39192 39910: contig of 719 bp in length
* 39911 40010: gap of 100 bp
* 40011 40740: contig of 730 bp in length
* 40741 40840: gap of 100 bp
* 40841 41571: contig of 731 bp in length
* 41572 41671: gap of 100 bp
* 41672 42431: contig of 760 bp in length
* 42432 42531: gap of 100 bp
* 42532 43249: contig of 718 bp in length
* 43250 43349: gap of 100 bp
* 43350 44088: contig of 739 bp in length
* 44089 44188: gap of 100 bp
* 44189 44950: contig of 762 bp in length
* 44951 45050: gap of 100 bp
* 45051 45813: contig of 763 bp in length
* 45814 45913: gap of 100 bp
* 45914 46609: contig of 696 bp in length
* 46610 46709: gap of 100 bp
```

```
* 46710 47423: contig of 714 bp in length
* 47424 47523: gap of 100 bp
* 47524 48208: contig of 685 bp in length
* 48209 48308: gap of 100 bp
* 48309 49010: contig of 702 bp in length
* 49011 49110: gap of 100 bp
* 49111 49889: contig of 779 bp in length
* 49890 49989: gap of 100 bp
* 49990 50577: contig of 588 bp in length
* 50578 50677: gap of 100 bp
* 50678 51404: contig of 727 bp in length
* 51405 51504: gap of 100 bp
* 51505 52236: contig of 732 bp in length
* 52237 52336: gap of 100 bp
* 52337 53095: contig of 759 bp in length
* 53096 53195: gap of 100 bp
* 53196 53943: contig of 748 bp in length
* 53944 54043: gap of 100 bp
* 54044 54802: contig of 759 bp in length
* 54803 54902: gap of 100 bp
* 54903 55630: contig of 728 bp in length
* 55631 55730: gap of 100 bp
* 55731 56410: contig of 680 bp in length
* 56411 56510: gap of 100 bp

Query Match 1.9%; Score 19; DB 78; Length 61435;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 211 gctgcagctcccaactc 229
DB 55825 GCTGCAGCTCCCACTTC 55807
```

Search completed: October 28, 2001, 22:55:01
Job time: 4014 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 21:51:47 ; Search time 102.68 Seconds

(without alignments)
6121.246 Million cell updates/sec

Title: SEQINS_COPY_1140_2140

Perfect score: 1001
Sequence: 1 tcgcagcccaagaattcaaa.....aagcattcagctgtgggatg 1001

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 8

Total number of hits satisfying chosen parameters: 731206

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	880	87.9	2546	21	Human ORFX ORF2000
2	880	87.9	2958	21	Human prostate can
3	727	72.6	2478	21	Human sulphatase G
4	124	12.4	2664	21	Human prostate can
5	24	2.4	24	21	Human prostate can
6	20	2.0	457	21	Human prostate can
7	19	1.9	19	21	Human HPC2 CDNA se
8	19	1.9	19	21	Human HPC2 CDNA se
9	18	1.8	18	21	Human HPC2 CDNA se
10	18	1.8	18	21	Human HPC2 CDNA se
11	18	1.8	18	21	Human HPC2 CDNA se

C	12	18	1.8	18	21	AAA60228	Human HPC2 CDNA se
C	13	18	1.8	18	21	AAA60237	Human HPC2 CDNA se
C	14	18	1.8	18	21	AAA60238	Human HPC2 CDNA se
C	15	18	1.8	18	21	AAA60240	Human HPC2 CDNA se
C	16	18	1.8	177	20	AAAX18066	Coding sequence fo
C	17	18	1.8	531	18	AAAT77453	H. pylori cytoplas
C	18	18	1.8	531	18	AAAT77453	H. pylori cytoplas
C	19	18	1.8	791	21	AAC44859	Arabidopsis thalia
C	20	18	1.8	1593	18	AAT67992	H. pylori cytoplas
C	21	18	1.8	1600	21	AAC77190	Human ORFX ORF2745
C	22	18	1.8	1626	20	AAAX08683	Novel nucleotide s
C	23	18	1.8	2186	19	AAV17351	Coding sequence fo
C	24	18	1.8	4120	20	AAZ09473	Human RNA helicase
C	25	17	1.7	17	21	AAA60239	Human HPC2 CDNA se
C	26	17	1.7	17	21	AAA60241	Human HPC2 CDNA se
C	27	17	1.7	17	21	AAA60242	Human HPC2 CDNA se
C	28	17	1.7	256	21	AAA32079	Plant microsatelli
C	29	17	1.7	292	21	AAAC32768	Arabidopsis thalia
C	30	17	1.7	300	21	AAA32050	Plant microsatelli
C	31	17	1.7	330	21	AAZ42643	Human 5' EST Isola
C	32	17	1.7	347	16	AAT24629	Human gene signatu
C	33	17	1.7	350	21	AAAC07468	Human secreted pro
C	34	17	1.7	398	22	AAAF64559	Novel human polynu
C	35	17	1.7	426	21	AAA32069	Plant microsatelli
C	36	17	1.7	431	21	AAA32052	Plant microsatelli
C	37	17	1.7	459	21	AAC46096	Arabidopsis thalia
C	38	17	1.7	523	21	AAC45096	Arabidopsis thalia
C	39	17	1.7	559	19	AAV29360	Calcium ion channe
C	40	17	1.7	694	22	AAAF64735	Novel human polynu
C	41	17	1.7	791	20	AAAX24979	Clone G156 encodi
C	42	17	1.7	855	21	AAZ91823	Streptococcus pneu
C	43	17	1.7	1022	20	AAAX40199	MAGE-4 encoding ge
C	44	17	1.7	1084	15	AAO67866	HE/MAGE-1 expressi
C	45	17	1.7	1084	20	AAZ08442	HE/MAGE-1 expressi

ALIGNMENTS

RESULT 1	
AACT6445	
ID	AACT6445 standard; CDNA; 2546 BP.
AC	AACT6445;
DT	08-FEB-2001 (first entry)
DE	Human ORFX ORF2000 polynucleotide sequence SEQ ID NO:3999.
XX	Human; open reading frame; ORFX; detection; cytosstatic; hepatotropic;
XX	vulnerary; antiporiatic; antiparkinsonian; noctropic; neuroprotective;
XX	anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
XX	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX	hypotensive; dermatological; immunosuppressive; antineoplastic; antihypertensive;
XX	antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;
XX	antianemic; gene therapy; cancer; proliferative disorder; hypertension;
XX	neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX	cholesterol ester storage; systemic lupus erythematosus; infection;
XX	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX	bone damage; cartilage damage; antineoplastic disease; coagulation;
XX	thrombosis; contraceptive; ss.
OS	Homo sapiens.
XX	WO200058473-A2.
XX	05-OCT-2000.
PD	31-MAR-2000; 2000WO-US08621.
XX	31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shmkets RA, leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB42236.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 3179-3180; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC osteoplastic; antiparkinsonian; nootropic; neuroprotective;
 CC antileptotic; anticonvulsant; antihypertensive; immunosuppressant;
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antitumoral; antineumatic;
 CC antihypertensive; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hyperextension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX
 SO Sequence 2546 BP; 652 A; 643 C; 686 G; 564 T; 1 other;

Query Match 87.9%; Score 880; DB 21; Length 2546;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1000; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 tgcagacacaaagatlaaaccagctcaactcatcatccaccgagacatctccctcgtc 60
 DB 738 tgcagacacaaagatlaaaccagctcaactcatcatccaccgagacatctccctcgtc 797
 QY 61 caccagttccgcgtgtaagaagagagccaccctcagtgcccatggttcagaggtga 120
 DB 798 caccagttccgcgtgtaagaagagagccaccctcagtggtcccatggttcagaggtga 857
 QY 121 atgctctcctaagtaacagctcgtgcccaagagaggtgagagaggtatccattatcac 180
 DB 858 atgctctcctaagtaacagctcgtgcccaagagaggtgagagaggtatccattatcac 917
 QY 181 ttgcaatccgaggaatcatcatggtgagagctgcagctcccaactccagcagagcgt 240
 DB 918 ttgcaatccgaggaatcatcatggtgagagctgcagctcccaactccagcagagcgt 977
 QY 241 gcaagagatcagagagaggtgagagagccagcccaacagaagaagaatcagta 300
 DB 978 gcaagagatcagagagaggtgagagagccagcccaacagaagaagaatcagta 1037
 QY 301 ccagaagaatcatctctcttggaacagaggtctcgcacccgaatgagaatcgaatgtcag 360
 DB 1038 ccagaagaatcatctctcttggaacagaggtctcgcacccgaatgagaatcgaatgtcag 1097
 QY 361 tgcacaactgttacaataagcccgacacagctctcgtactgactgactgtgtgagggcac 420
 DB 1098 tgcacaactgttacaataagcccgacacagctctcgtactgactgactgtgtgagggcac 1157

QY 421 attggcagctgtgcccattacagagaccaggtgacagggctcctggcacccctggc 480
 DB 1158 attggcagctgtgcccattacagagaccaggtgacagggctcctggcacccctggc 1217
 QY 481 tgcgtgtgtgtgtcccaactgagcagcagatcaccaacacagggcttgcgaatctcgc 540
 DB 1218 tgcgtgtgtgtgtcccaactgagcagcagatcaccaacacagggcttgcgaatctcgc 1276
 QY 541 tgcagagagaagcgcgccttgccatcttgggaagccgcctcacccttggctgtgttg 600
 DB 1277 tgcagagagaagcgcgccttgccatcttgggaagccgcctcacccttggctgtgttg 1336
 QY 601 cccccaacagctcaaaagccttgctccagcagtaaccacaacaggtgcagaaggtctcgc 660
 DB 1337 cccccaacagctcaaaagccttgctccagcagtaaccacaacaggtgcagaaggtctcgc 1396
 QY 661 accaatcatgatatcttcctcgaatgctccttcaggaagggcgtgagatctcagctcgt 720
 DB 1397 accaatcatgatatcttcctcgaatgctccttcaggaagggcgtgagatctcagctcgt 1456
 QY 721 cagtggaagatgatcatgctgctgttggacacatgatttgaagagttcagaact 780
 DB 1457 cagtggaagatgatcatgctgctgttggacacatgatttgaagagttcagaact 1516
 QY 781 gtcgtgtgtgtgtcccaactgagcagcagatcgttgcgtgtgcgtgtgcacacctgtcga 840
 DB 1517 gtcgtgtgtgtgtcccaactgagcagcagatcgttgcgtgtgcgtgtgcacacctgtcga 1576
 QY 841 aagtgtctatctccgggagacacatgcctcgtgagagctgtgtccgatatgggaagatg 900
 DB 1577 aagtgtctatctccgggagacacatgcctcgtgagagctgtgtccgatatgggaagatg 1636
 QY 901 ccaccctccttatcatagaagccaccccttggaagatgttgggaagagaacatgtgaaa 960
 DB 1637 ccaccctccttatcatagaagccaccccttggaagatgttgggaagagaacatgtgaaa 1696
 QY 961 agacacacagcacaacgctcccaagccatcagcgtgagagatg 1001
 DB 1697 agacacacagcacaacgctcccaagccatcagcgtgagagatg 1737

RESULT 2
 AAA58453
 ID AAA58453 standard; cDNA; 2958 BP.
 XX
 AC AAA58453;
 XX
 DT 07-DEC-2000 (first entry)
 XX
 DE Human prostate cancer predisposing gene HPC2 coding sequence.
 XX
 KW Human: prostate cancer predisposing gene; HPC2; chromosome 17p;
 KW gene therapy; peptide therapy; drug design; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT CDS 51..2531
 FT /*tag= a
 FT /product= "HPC2"
 PN W0200027864-A1.
 PD 18-MAY-2000.
 XX
 PF 05-NOV-1999; 99MO-US26055.
 XX
 PR 06-NOV-1998; 98US-0107468.
 XX
 PA (MYRI-) MYRIAD GENETICS INC.
 XX
 PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;

```
XX MPI: 2000-376481/32.
DR P-PSDB; AAB07228.
XX Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer
XX
PS Claim 3: Page 98-100; 157pp; English.
XX
CC The present sequence is the coding sequence of the human prostate
CC cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC alleles of this gene cause a predisposition to cancer, particularly
CC prostate cancer. This gene and its protein can be used in peptide and
CC gene therapy for cancer patients, as well as being useful as diagnostic
CC tools (both for cancer sufferers and those with a predisposition to the
CC disease) and in the production of cancer drugs. This sequence was
CC isolated by cloning and sequencing the region of the genome which
CC appeared to cause a predisposition to prostate cancer.
XX
SQ Sequence 2958 BP; 707 A; 805 C; 848 G; 598 T; 0 other;

Query Match      87.9%; Score 880; DB 21; Length 2958;
Best local Similarity 99.9%; Pred. No. 0;
Matches 1000; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 tcgcagccacagatccaacacagctcaacctcaccgcggacatcttccctgct 60
   |||||||
DB 1190 tcgcagccacagatccaacacagctcaacctcaccgcggacatcttccctgct 1249
   |||||||

QY 61 caccagtttcgctgttaagaagaagagcccccaccctcagctgctcccttgaaggtga 120
   |||||||
DB 1250 caccagtttcgctgttaagaagaagagcccccaccctcagctgctcccttgaaggtga 1309
   |||||||

QY 121 atgcctcctcaagttaccagctccctcccaaggagagtggtgacagaagatccattatc 180
   |||||||
DB 1310 atgcctcctcaagttaccagctccctcccaaggagagtggtgacagaagatccattatc 1369
   |||||||

QY 181 ttgcattcctggaagaattcattagttgagcgctgcagcttcccaacttccagcagagcgt 240
   |||||||
DB 1370 ttgcattcctggaagaattcattagttgagcgctgcagcttcccaacttccagcagagcgt 1429
   |||||||

QY 241 gcagaggtacagagagagtgacgacagagcccgccagcccgagagagaagaagatcagta 300
   |||||||
DB 1430 gcagaggtacagagagagtgacgacagagcccgccagcccgagagagaagaagatcagta 1489
   |||||||

QY 301 cccaagaatcattctcttggaaacaggtctgcattcccgatgaagaattcgaatgtcag 360
   |||||||
DB 1490 cccaagaatcattctcttggaaacaggtctgcattcccgatgaagaattcgaatgtcag 1549
   |||||||

QY 361 tgcacaacttgttaacataaagcccgacacgtctctgtctagtgtgtgagggcac 420
   |||||||
DB 1550 tgcacaacttgttaacataaagcccgacacgtctctgtctagtgtgtgagggcac 1609
   |||||||

QY 421 atttgggcaagctgtgcgctattacgagagacaggttgaaaggtctctgggcaaccctggc 480
   |||||||
DB 1610 atttgggcaagctgtgcgctattacgagagacaggttgaaaggtctctgggcaaccctggc 1669
   |||||||

QY 481 tgcgtgtgtgtgttcccaacctgtgcacgacatcacccacagggcttgcagaatattctgc 540
   |||||||
DB 1670 tgcgtgtgtgtgttcccaacct-gcagcgagatcacccacagggcttgcagaatattctgc 1728
   |||||||

QY 541 tgcagagagagacgagcgttggcatcttgggaagcgcttcaaccttgcgtgtgttg 600
   |||||||
DB 1729 tgcagagagagacgagcgttggcatcttgggaagcgcttcaaccttgcgtgtgttg 1788
   |||||||

QY 601 cccccaacagctccaagcctgtgctccagcagatcaccaacacaggtgcagaagagctctgc 660
   |||||||
DB 1789 cccccaacagctccaagcctgtgctccagcagatcaccaacacaggtgcagaagagctctgc 1848
   |||||||

QY 661 accacatcagtatgatctctgccaatgaccttcagaagaagggtcagatctcagctctg 720
   |||||||
DB 1849 accacatcagtatgatctctgccaatgaccttcagaagaagggtcagatctcagctctg 1908
   |||||||
```

```
QY 721 cagtggaagaatgatcatcagctcgttggcgaacatgtgatttggaaagagttccagacct 780
   |||||||
DB 1909 cagtggaagaatgatcatcagctcgttggcgaacatgtgatttggaaagagttccagacct 1968
   |||||||

QY 781 gctgtgtgagcagctgcaacatgcttggctgtgtgcgctgtgcaacctctgcgcgga 840
   |||||||
DB 1969 gctgtgtgagcagctgcaacatgcttggctgtgtgcgctgtgcaacctctgcgcgga 2028
   |||||||

QY 841 aagtggtctatctccgggggaacacatgcccctgcagagctctgttcggatgtgggaagatg 900
   |||||||
DB 2029 aagtggtctatctccgggggaacacatgcccctgcagagctctgttcggatgtgggaagatg 2088
   |||||||

QY 901 ccacccctccgatacatatgaagccacccctggaagaatgttgggaagagagcagctgga 960
   |||||||
DB 2089 ccacccctccgatacatatgaagccacccctggaagaatgttgggaagagagcagctgga 2148
   |||||||

QY 961 agacacacagcacacagctcccaagccatcagctgtgggagtg 1001
   |||||||
DB 2149 agacacacagcacacagctcccaagccatcagctgtgggagtg 2189
   |||||||

RESULT 3
AAA52810
ID AAA52810 standard; cDNA; 2478 BP.
XX
XX AAA52810;
AC
XX
XX 20-SEP-2000 (first entry)
DT
XX
XX Human sulphatase G cDNA.
DE
XX Human sulphatase G; hSG; chromosome 17p11.2; gene therapy; ss.
KM
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 1.2478
FT CDS
FT /tag- a
FT /partial
FT /product- "hSG"
XX
XX MO200034327-A1.
XX
XX 15-JUN-2000.
PD
XX
XX 09-DEC-1999; 99WC-AU01092.
PF
XX
XX 09-DEC-1998; 98AU-0007624.
PR
XX
XX (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
PA
XX Hopwood JJ, Liltjens T, Hu RL;
XX
XX
XX MPI: 2000-431273/37.
DR P-PSDB; AAY9850.
XX
XX Novel isolated DNA sequence which encodes human sulfatase G or its
PT fragment useful in gene therapy for treating patients suffering from
PT sulfatase deficiency -
XX
PS Claim 2: Page 29-30; 33pp; English.
XX
XX The present sequence encodes human sulphatase G (hSG). hSG is
XX CC not a member of the well-characterised CYP5R sulphatase family.
XX CC It belongs to a family showing sequence similarity to a sulphatase
XX CC from the marine bacterium Alteromonas carraglenovora. The hSG gene
XX CC contains 23 exons and is located at chromosome 17p11.2. The present
XX CC sequence is clone lambda29.1 from a human testes cDNA library. It was
XX CC isolated using human EST sequences with sequence similarity to the
XX CC non-CYP5R family as a probe to screen the library. The cDNA insert was
XX CC subcloned and the DNA sequence of both strands was determined. The
XX CC sequence may be used to treat a patient suffering from hSG deficiency
```

CC by replacing, repairing, or compensating for a DNA sequence within that
CC patient's genome.
XX
SQ Sequence 2478 BP; 587 A; 686 C; 709 G; 496 T; 0 other:

Query Match 72.6%; Score 727; DB 21; Length 2478;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 997; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 tgcgagccacgaattcaaccacagctcaacctcaatccacccgaatattccctctgt 60
DB 1140 tgcgagccacgaattcaaccacagctcaacctcaatccacccgaatattccctctgt 1199
QY 61 caccgaattccgctgtgaagaagagggcccaacctcagttgtccatggttcagggtta 120
DB 1200 caccgaattccgctgtgaagaagagggcccaacctcagttgtccatggttcagggtta 1259
QY 121 atgctctctcaagttacagctccgctccaggaaggaagtgagcagaaggaatgcatat 180
DB 1260 atgctctctcaagttacagctccgctccaggaaggaagtgagcagaaggaatgcatat 1319
QY 181 tgcgaatccttgaggaattcatagttgagcgctgtcagcttcccaattccagcagagct 240
DB 1320 tgcgaatccttgaggaattcatagttgagcgctgtcagcttcccaattccagcagagcgt 1379
QY 241 gcagaggtatacagaagaggtgcgcagagcgcccgagcccgagcagagaagaaagatcagta 300
DB 1380 gcagaggtatacagaagaggtgcgcagagcgcccgagcccgagcagagaagaaagatcagta 1439
QY 301 cccagaataatcatcttcccttggaacagggtctgcacatcccgatgaagattcgaatgtcag 360
DB 1440 cccagaataatcatcttcccttggaacagggtctgcacatcccgatgaagattcgaatgtcag 1499
QY 361 tgcacaacttgcacaataagcccgcagcgtctcgtcactcgtgactgtgtgagggcac 420
DB 1500 tgcacaacttgcacaataagcccgcagcgtctcgtcactcgtgactgtgtgagggcac 1559
QY 421 atttgagcgctgtgctgcattacgagagcaggtgagcagaggtcttgagcaccctggc 480
DB 1560 atttgagcgctgtgctgcattacgagagcaggtgagcagaggtcttgagcaccctggc 1619
QY 481 tgcgtgttctgtgtcccaactgycacgagatcaaccacacgggctgtccaaagtatctgc 540
DB 1620 tgcgtgttctgtgtcccaactgycacgagatcaaccacacgggctgtccaaagtatctgc 1678
QY 541 tgcagagagaagcgcgcttgcacatcttgggaagcgcgcttcaaccttgcgtgtgttg 600
DB 1679 tgcagagagaagcgcgcttgcacatcttgggaagcgcgcttcaaccttgcgtgtgttg 1738
QY 601 cccccaacagctcaaaagctgtgctcagcagatcaacaacacagttccagagggtctctgc 660
DB 1739 cccccaacagctcaaaagctgtgctcagcagatcaacaacacagttccagagggtctctgc 1798
QY 661 accacaatcagtatgtatccctgcgaatgtccttcagaagaagggctgagatctccagtcctg 720
DB 1799 accacaatcagtatgtatccctgcgaatgtccttcagaagaagggctgagatctccagtcctg 1858
QY 721 cagtggaagaatgtatcagatctgcgtgttggaacatgtatatttggaagatttcagacct 780
DB 1859 cagtggaagaatgtatcagatctgcgtgttggaacatgtatatttggaagatttcagacct 1918
QY 781 gtcgtgtgagcagtcagcagatgcgtgttggtgctgtgcgtgtgtgacacactctggtctga 840
DB 1919 gtcgtgtgagcagtcagcagatgcgtgttggtgctgtgcgtgtgtgacacactctggtctga 1978
QY 841 aagtggtctatctcgggggacacacatgccttcggaaggtcttggtccggatcggggaagatg 900
DB 1979 aagtggtctatctcgggggacacacatgccttcggaaggtcttggtccggatcggggaagatg 2038
QY 901 ccacctctctgatacagagccaccttggaagatggtttggaaggaagcagtcagtaaa 960
DB 2039 ccacctctctgatacagagccaccttggaagatggtttggaaggaagcagtcagtaaa 2098

QY 961 agacacacagcacacgctcccaagccatcagtcgttgggagtg 1001
DB 2099 agacacacagcacacgctcccaagccatcagtcgttgggagtg 2139

RESULT 4
AAA60207
ID AAA60207 standard; DNA; 26664 BP.

AC AAA60207;

DT 07-DEC-2000 (first entry)

DE Human prostate cancer predisposing gene HPC2 genomic sequence.

KW Human; prostate cancer predisposing gene; HPC2; chromosome 17p;

KW gene therapy; peptide therapy; drug design; ds.

OS Homo sapiens.

FT Key
FT CDS
FT location/Qualifiers

FT 910..26039

FT /*tag= a

FT /*product= "HPC2"

FT /*note= "this sequence contains introns"

FT /*transl_except= (pos:23892..23895,aa:glu)

FT 910..1154

FT /*tag= b

FT /*number= 1

FT 1736..1786

FT /*tag= c

FT /*number= 2

FT 1925..1995

FT /*tag= d

FT /*number= 3

FT 3025..3089

FT /*tag= e

FT /*number= 4

FT 4361..4418

FT /*tag= f

FT /*number= 5

FT 5582..5650

FT /*tag= g

FT /*number= 6

FT 7075..7194

FT /*tag= h

FT /*number= 7

FT 8186..8244

FT /*tag= i

FT /*number= 8

FT 12878..12936

FT /*tag= j

FT /*number= 9

FT 13032..13104

FT /*tag= k

FT /*number= 10

FT 13756..13868

FT /*tag= l

FT /*number= 11

FT 15283..15378

FT /*tag= m

FT /*number= 12

FT 16278..16416

FT /*tag= n

FT /*number= 13

FT 16498..16583

FT /*tag= o

FT /*number= 14

FT 18583..18701

FT /*tag= p

FT /*number= 15

FT 20349..20445

FT exon

FT exon

```

FT      /tag= q
FT      /number= 16
FT      22172..22310
FT      /tag= t
FT      /number= 17
FT      22879..22917
FT      /tag= s
FT      /number= 18
FT      23045..23154
FT      /tag= t
FT      /number= 19
FT      23795..23895
FT      /tag= u
FT      /number= 20
FT      23973..24093
FT      /tag= v
FT      /number= 21
FT      24354..24432
FT      /tag= w
FT      /number= 22
FT      25026..25170
FT      /tag= x
FT      /number= 23
FT      25812..26036
FT      /tag= y
FT      /number= 24
FT      26447..26452
FT      /tag= z
FT      WO200027864-A1.
XX      18-MAY-2000.
XX      PD
XX      PF 05-NOV-1999; 99WO-US26055.
XX      PR 06-NOV-1998; 98US-0107468.
XX      PA (MYRI-) MYRIAD GENETICS INC.
XX      PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX      DR WPI; 2000-376481/32.
XX      P-PSDB; AAB07228.
XX      PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
XX      antibodies, useful for treatment and diagnosis of prostate cancer
XX      Claim 3; Page 108-122; 157pp; English.
XX      CC The present sequence is the genomic sequence of the human prostate
XX      cancer predisposing gene HPC2, which is found on chromosome 17p. Some
XX      alleles of this gene cause a predisposition to cancer, particularly
XX      prostate cancer. This gene and its protein can be used in peptide and
XX      gene therapy for cancer patients, as well as being useful as diagnostic
XX      tools (both for cancer sufferers and those with a predisposition to the
XX      disease) and in the production of cancer drugs. This sequence was
XX      isolated by cloning and sequencing the region of the genome which
XX      appeared to cause a predisposition to prostate cancer.
XX      SQ Sequence 26664 BP; 6173 A; 6300 C; 6519 G; 7661 T; 11 other;

Query Match      12.4%; Score 124; DB 21; Length 26664;
Best Local Similarity 100.0%; Pred. No. 1.4e-52;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 agtttcagacctctgctgctgcgcacactgcaagatgcttgcctgctgctggtgcaca 828
DB 23971 agtttcagacctctgctgctgcgcacactgcaagatgcttgcctgctgctggtgcaca 24030

QY 829 cctctgctggaagtgtctatctccgggacacacatgcccctcggaagctgtgctcgga 888
DB 24031 cctctgctggaagtgtctatctccgggacacacatgcccctcggaagctgtgctcgga 24090

```

```

QY 889 tggg 892
DB 24091 tggg 24094

RESULT 5
ID AAA60374 standard; cDNA; 24 BP.
XX AC AAA60374;
XX AC AAA60374;
XX DT 07-DEC-2000 (first entry)
XX DE Human prostate cancer predisposing gene HPC2 variant 1641insG.
XX KW Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
XX KM gene therapy; peptide therapy; drug design; ss.
XX OS Homo sapiens.
XX PN WO200027864-A1.
XX PD 18-MAY-2000.
XX PF 05-NOV-1999; 99WO-US26055.
XX PR 06-NOV-1998; 98US-0107468.
XX PA (MYRI-) MYRIAD GENETICS INC.
XX PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX DR WPI; 2000-376481/32.
XX PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
XX PT antibodies, useful for treatment and diagnosis of prostate cancer
XX PS Example 5; Page 63; 157pp; English.
XX CC The present sequence is part of the coding sequence of a variant of the
XX CC human prostate cancer predisposing gene HPC2, which is found on
XX CC chromosome 17p. This variant has been designated 1641insG. Some
XX CC alleles of this gene cause a predisposition to cancer, particularly
XX CC prostate cancer, this allele being an example of this. The HPC2 gene and
XX CC its protein can be used in peptide and gene therapy for cancer patients,
XX CC as well as being useful as diagnostic tools (both for cancer sufferers
XX CC and those with a predisposition to the disease) and in the production of
XX CC cancer drugs. This sequence was isolated by mutation screening of the
XX CC HPC2 gene in humans.
XX SQ Sequence 24 BP; 5 A; 9 C; 6 G; 4 T; 0 other;

Query Match      2.4%; Score 24; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 491 gtgtccacctgacgcagatca 514
DB 1 gtgtccacctgacgcagatca 24

RESULT 6
ID AAC57035 standard; DNA; 457 BP.
XX AC AAC57035;
XX AC AAC57035;
XX DT 25-JAN-2001 (first entry)
XX DE Plusa radiata transcription factor DNA sequence #481.
XX

```

KW plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bzip; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox;
KW homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;
KW type 2 CysHis2; CCAAT box element; MYB; ss.
XX
OS Pinus radiata.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
PI WPI, 2000-579369/54.
XX
DR
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT -
XX
PS Claim 1; Pages 570-571; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bzip, bzip family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and ERBS, zinc finger domains of type 2 CysHis2, CCAAT box elements
CC and MYB.
XX
SQ Sequence 457 BP; 130 A; 94 C; 118 G; 115 T; 0 other;

Query Match 2.0%; Score 20; DB 21; Length 457;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 944 agaggaagcagtcggaaga 963
|||||
DB 146 agaggaagcagtcggaaga 165

RESULT 7
AAA60229
ID AAA60229 standard; DNA; 19 BP.
XX
AC AAA60229;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 50.
XX
KW Human; mouse; prostate cancer predisposing gene; HPC2;
KW human chromosome 17p; gene therapy; peptide therapy; drug design;
KW PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX
PN WO200027864-A1.

XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
PI WPI, 2000-376481/32.
XX
DR
XX
PT Human prostate cancer (HPC2) nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer
PT -
XX
PS Example 3; Page 55; 157pp; English.
XX
CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 19 BP; 5 A; 3 C; 6 G; 5 T; 0 other;

Query Match 1.9%; Score 19; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 763 tggagagtttcagacctg 781
|||||
DB 1 tggagagtttcagacctg 19

RESULT 8
AAA60230/C
ID AAA60230 standard; DNA; 19 BP.
XX
AC AAA60230;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 51.
XX
KW Human; mouse; prostate cancer predisposing gene; HPC2;
KW human chromosome 17p; gene therapy; peptide therapy; drug design;
KW PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
PI WPI, 2000-376481/32.
XX
DR
XX
PT Human prostate cancer (HPC2) nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer
PT -
XX
PS Example 3; Page 55; 157pp; English.

XX The present sequence is a primer used in the isolation of the human
 CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
 CC version of the gene is found on chromosome 17p. Some alleles cause a
 CC predisposition to cancer, particularly prostate cancer. This gene and its
 CC protein can be used in peptide and gene therapy for cancer patients, as
 CC well as being useful as diagnostic tools (both for cancer sufferers and
 CC those with a predisposition to the disease) and in the production of
 CC cancer drugs.

XX Sequence 19 BP; 5 A; 6 C; 3 G; 5 T; 0 other;

Query Match 1.8%; Score 19; DB 21; Length 19;

Best Local Similarity 100.0%; Pred. No. 9.8; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 tgaagagcttcacagctg 781

Db 19 TGGAGAGCTTCAGACCTG 1

RESULT 9

AAA60225

ID AAA60225 standard; DNA; 18 BP.

XX AAA60225;

DT 07-DEC-2000 (first entry)

DE Human HPC2 cDNA sequencing primer SEQ ID NO: 46.

KM Human; mouse; prostate cancer predisposing gene; HPC2;

KW human chromosome 17p; gene therapy; peptide therapy; drug design;

KM PCR primer; sequencing primer; ss.

OS Homo sapiens.

XX WO200027864-A1.

XX 18-MAY-2000.

PF 05-NOV-1999; 99WO-US26055.

PR 06-NOV-1998; 98US-0107468.

PA (MYRI-) MYRIAD GENETICS INC.

PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;

DR WPI; 2000-376481/32.

PT Human prostate cancer (HPC2) nucleic acids, polypeptides, and

XX antibodies, useful for treatment and diagnosis of prostate cancer -

PS Example 3; Page 55; 157pp; English.

XX The present sequence is a primer used in the isolation of the human
 CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
 CC version of the gene is found on chromosome 17p. Some alleles cause a
 CC predisposition to cancer, particularly prostate cancer. This gene and its
 CC protein can be used in peptide and gene therapy for cancer patients, as
 CC well as being useful as diagnostic tools (both for cancer sufferers and
 CC those with a predisposition to the disease) and in the production of
 CC cancer drugs.

XX Sequence 18 BP; 5 A; 5 C; 3 G; 5 T; 0 other;

Query Match 1.8%; Score 18; DB 21; Length 18;

Best Local Similarity 100.0%; Pred. No. 31; Mismatches 0; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 tgaatgctctcctaagta 135

Db 1 tgaatgctctcctaagta 18

RESULT 10

AAA60226/c

ID AAA60226 standard; DNA; 18 BP.

XX AAA60226;

DT 07-DEC-2000 (first entry)

DE Human HPC2 cDNA sequencing primer SEQ ID NO: 47.

KM Human; mouse; prostate cancer predisposing gene; HPC2;

KW human chromosome 17p; gene therapy; peptide therapy; drug design;

KM PCR primer; sequencing primer; ss.

OS Homo sapiens.

XX WO200027864-A1.

XX 18-MAY-2000.

PF 05-NOV-1999; 99WO-US26055.

PR 06-NOV-1998; 98US-0107468.

PA (MYRI-) MYRIAD GENETICS INC.

PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;

DR WPI; 2000-376481/32.

PT Human prostate cancer (HPC2) nucleic acids, polypeptides, and

XX antibodies, useful for treatment and diagnosis of prostate cancer -

PS Example 3; Page 55; 157pp; English.

XX The present sequence is a primer used in the isolation of the human
 CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
 CC version of the gene is found on chromosome 17p. Some alleles cause a
 CC predisposition to cancer, particularly prostate cancer. This gene and its
 CC protein can be used in peptide and gene therapy for cancer patients, as
 CC well as being useful as diagnostic tools (both for cancer sufferers and
 CC those with a predisposition to the disease) and in the production of
 CC cancer drugs.

XX Sequence 18 BP; 5 A; 3 C; 5 G; 5 T; 0 other;

Query Match 1.8%; Score 18; DB 21; Length 18;

Best Local Similarity 100.0%; Pred. No. 31; Mismatches 0; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 tgaatgctctcctaagta 135

Db 18 TGAATGCTCTCTCAAGTA 1

RESULT 11

AAA60227

ID AAA60227 standard; DNA; 18 BP.

XX AAA60227;

DT 07-DEC-2000 (first entry)

DE Human HPC2 cDNA sequencing primer SEQ ID NO: 48.

KM Human; mouse; prostate cancer predisposing gene; HPC2;

KW human chromosome 17p; gene therapy; peptide therapy; drug design;

KW PCR primer; sequencing primer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200027864-A1.
 XX
 PD 18-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26055.
 XX
 PR 06-NOV-1998; 98US-0107468.
 XX
 PA (MYRI-) MYRIAD GENETICS INC.
 XX
 PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
 XX WPI: 2000-376481/32.
 DR
 XX
 PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
 XX antibodies, useful for treatment and diagnosis of prostate cancer -
 XX
 PS Example 3; Page 55; 157pp; English.
 XX
 CC The present sequence is a primer used in the isolation of the human
 CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
 CC version of the gene is found on chromosome 17p. Some alleles cause a
 CC predisposition to cancer, particularly prostate cancer. This gene and its
 CC protein can be used in peptide and gene therapy for cancer patients, as
 CC well as being useful as diagnostic tools (both for cancer sufferers and
 CC those with a predisposition to the disease) and in the production of
 CC cancer drugs.
 CC
 XX
 SQ Sequence 18 BP; 3 A; 3 C; 7 G; 5 T; 0 other;

Query Match 1.8%; Score 18; DB 21; Length 18;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 397 gctactgactgtgtga 414
 ||||||||||||||||
 Db 1 gctactgactgtgtga 18

RESULT 12
 AAA60228/C
 ID AAA60228 standard; DNA; 18 BP.
 XX
 AC AAA60228;
 XX
 DT 07-DEC-2000 (first entry)
 XX
 DE Human HPC2 cDNA sequencing primer SEQ ID NO: 49.
 XX
 KW Human; mouse; prostate cancer predisposing gene; HPC2;
 KW human chromosome 17p; gene therapy; peptide therapy; drug design;
 KW PCR primer; sequencing primer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200027864-A1.
 XX
 PD 18-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26055.
 XX
 PR 06-NOV-1998; 98US-0107468.
 XX
 PA (MYRI-) MYRIAD GENETICS INC.
 XX
 PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
 XX WPI: 2000-376481/32.
 DR

XX
 PR Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
 XX antibodies, useful for treatment and diagnosis of prostate cancer -
 XX
 PS Example 3; Page 55; 157pp; English.
 XX
 CC The present sequence is a primer used in the isolation of the human
 CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
 CC version of the gene is found on chromosome 17p. Some alleles cause a
 CC predisposition to cancer, particularly prostate cancer. This gene and its
 CC protein can be used in peptide and gene therapy for cancer patients, as
 CC well as being useful as diagnostic tools (both for cancer sufferers and
 CC those with a predisposition to the disease) and in the production of
 CC cancer drugs.
 CC
 XX
 SQ Sequence 18 BP; 5 A; 7 C; 3 G; 3 T; 0 other;

Query Match 1.8%; Score 18; DB 21; Length 18;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 397 gctactgactgtgtga 414
 ||||||||||||||||
 Db 18 GCTACTGACTGTGTGA 1

RESULT 13
 AAA60237
 ID AAA60237 standard; DNA; 18 BP.
 XX
 AC AAA60237;
 XX
 DT 07-DEC-2000 (first entry)
 XX
 DE Human HPC2 cDNA sequencing primer SEQ ID NO: 58.
 XX
 KW Human; mouse; prostate cancer predisposing gene; HPC2;
 KW human chromosome 17p; gene therapy; peptide therapy; drug design;
 KW PCR primer; sequencing primer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200027864-A1.
 XX
 PD 18-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26055.
 XX
 PR 06-NOV-1998; 98US-0107468.
 XX
 PA (MYRI-) MYRIAD GENETICS INC.
 XX
 PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
 XX WPI: 2000-376481/32.
 DR
 XX
 PR Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
 XX antibodies, useful for treatment and diagnosis of prostate cancer -
 XX
 PS Example 3; Page 56; 157pp; English.
 XX
 CC The present sequence is a primer used in the isolation of the human
 CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
 CC version of the gene is found on chromosome 17p. Some alleles cause a
 CC predisposition to cancer, particularly prostate cancer. This gene and its
 CC protein can be used in peptide and gene therapy for cancer patients, as
 CC well as being useful as diagnostic tools (both for cancer sufferers and
 CC those with a predisposition to the disease) and in the production of
 CC cancer drugs.
 CC
 XX
 SQ Sequence 18 BP; 6 A; 7 C; 3 G; 2 T; 0 other;

Query Match	1.8%;	Score 18;	DB 21;	Length 18;
Best Local Similarity	100.0%;	Pred. No. 31;		
Matches 18;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	371	gtcacataagcccgac	388	
DB	1	gtcacataagcccgac	18	
RESULT 14				
ID	AAA60238	standard; DNA; 18 BP.		
XX	AAA60238;			
XX	07-DEC-2000	(first entry)		
DE	Human HPC2 cDNA sequencing primer SEQ ID NO: 59.			
XX				
KW	Human; mouse; prostate cancer predisposing gene; HPC2;			
KM	human chromosome 17p; gene therapy; peptide therapy; drug design;			
KW	PCR primer; sequencing primer; ss.			
OS	Homo sapiens.			
XX				
PN	WO200027864-A1.			
XX				
PD	18-MAY-2000.			
XX				
PF	05-NOV-1999;	99WO-US26055.		
XX				
PR	06-NOV-1998;	98US-0107468.		
XX				
PA	(MYRI-) MYRIAD GENETICS INC.			
XX				
PI	Tavligian SV, Teng DHF, Simard J, Rommens JW;			
XX				
DR	WPI: 2000-376481/32.			
XX				
PT	Human prostate cancer (HPC2) nucleic acids, polypeptides, and			
XX	antibodies, useful for treatment and diagnosis of prostate cancer -			
PS	Example 3: Page 56; 157pp; English.			
XX				
CC	The present sequence is a primer used in the isolation of the human			
CC	and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human			
CC	version of the gene is found on chromosome 17p. Some alleles cause a			
CC	predisposition to cancer, particularly prostate cancer. This gene and its			
CC	protein can be used in peptide and gene therapy for cancer patients, as			
CC	well as being useful as diagnostic tools (both for cancer sufferers and			
CC	those with a predisposition to the disease) and in the production of			
CC	cancer drugs.			
XX				
SO	Sequence 18 BP; 0 A; 3 C; 7 G; 8 T; 0 other;			
Query Match 1.8%; Score 18; DB 21; Length 18;				
ID	AAA60240	standard; DNA; 18 BP.		
XX	AAA60240;			
XX	07-DEC-2000	(first entry)		

XX		Human HPC2 cDNA sequencing primer SEQ ID NO: 61.
DE		
XX		Human; mouse; prostate cancer predisposing gene; HPC2;
KW		human chromosome 17p; gene therapy; peptide therapy; drug design;
KM		PCR primer; sequencing primer; ss.
XX		
OS	Homo sapiens.	
XX		
PX	WO200027864-A1.	
PN		
XX	18-MAY-2000.	
PD		
XX		
PF	05-NOV-1999;	99WO-US26055.
PR		
XX	06-NOV-1998;	98US-0107468.
XX		
PA	(MYRI-) MYRIAD GENETICS INC.	
PI	Tavtigian SV, Teng DHF, Simard J, Rommens JM;	
DR		
XX	WPI; 2000-376481/32.	
XX		
PT	Human prostate cancer (HPC)2 nucleic acids, polypeptides, and	
PT	antibodies, useful for treatment and diagnosis of prostate cancer -	
PS	Example 3; Page 56; 157pp; English.	
XX		
CC	The present sequence is a primer used in the isolation of the human	
CC	and murine prostate cancer predisposing genes HPC2 and Km.HPC2. The human	
CC	version of the gene is found on chromosome 17p. Some alleles cause a	
CC	predisposition to cancer, particularly prostate cancer. This gene and its	
CC	protein can be used in peptide and gene therapy for cancer patients, as	
CC	well as being useful as diagnostic tools (both for cancer sufferers and	
CC	those with a predisposition to the disease) and in the production of	
CC	cancer drugs.	
SX	Sequence 18 BP; 5 A; 4 C; 3 G; 6 T; 0 other;	
QY	Query Match	1.8%; Score 18; DB 21; Length 18;
	Best Local Similarity	100.0%; Pred. No. 31;
	Matches 18; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	671 tatgattcctgcgaatg 688	
Db	1 tatgattcctgcgaatg 18	
RESULT 16		
AAI18066/C		
ID	AAI18066 standard; DNA; 177 BP.	
XX		
AC	AAI18066;	
XX		
DT	04-MAY-1999 (first entry)	
XX		
DE	Coding sequence for human SI binding protein SNI45.	
KW	Gastro-intestinal transport receptor; binding protein; hSI; HP11;	
KM	D2H; hPEP1; human; GI tract receptor; sucrose-isomaltase complex;	
KW	intestinal peptide-associated transporter; hypertension; diabetes;	
KM	osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;	
KM	therapeutic agent delivery; therapy; ss.	
OS	Homo sapiens.	
XX		
PN	WO9851325-A2.	
XX		
PD	19-NOV-1998.	
XX		
PF	15-MAY-1998;	98WO-US10088.
XX		

PR 15-MAY-1997; 97US-0046595.
XX
XX (CYTO-) CYTOGEN CORP.
PA (ELAN-) ELAN CORP PLC.
XX
XX Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JF;
PI Omahony DJ, Patterson CA, Singleton J;
XX
XX WPI: 1999-009568/01.
DR
XX
XX New proteins that bind specifically to receptors in the
PT gastro-intestinal tract and related nucleic acid - chimaeras and
PT antibodies, used to deliver therapeutic or diagnostic agents to, or
PT through, the gastrointestinal tract, e.g. insulin or leuprolide
XX
XX
PS Claim 49; Page 56; 294pp: English.
XX
XX This sequence encodes a peptide that specifically binds to the human
CC sucrose-isomaltase complex. The invention relates to purified
CC proteins (I) that bind specifically to at least one of the
CC gastro-intestinal (GI) tract receptors human intestinal
CC peptide-associated transporter (HPII), hPEPT1, D2H and human
CC sucrose-isomaltase complex (HSI). (I) provide active transport of
CC therapeutic agents through human and animal GI tissue (into the blood)
CC for in vivo delivery, particularly for treatment or prevention of
CC hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,
CC migraine, or angina pectoris. Specifically they are used to deliver
CC insulin or leuprolide, but many other suitable therapeutic agents are
CC disclosed, including genes or inhibitory nucleic acids, imaging agents and
CC antigens. (I) may also provide targeting to the GI tract. Other uses of
CC (I) are: (i) to determine the level of specified receptors in a sample
CC (in a binding assay); and (ii) to screen for molecules that bind (I).
CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC and in immunoassays. The antibodies are used to locate, detect and
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.
XX
XX Sequence 177 BP; 31 A; 51 C; 53 G; 42 T; 0 other;
SQ
XX
XX Query Match 1.8%; Score 18; DB 20; Length 177;
XX Best Local Similarity 100.0%; Pred. No. 30;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 628 agcagtcacacacagc 645
DB 90 AGCAGTACCCACACCACT 73
XX
XX
XX RESULT 17
XX AAT67773/C
XX ID AAT67773 standard; DNA; 531 BP.
XX
XX AAT67773;
XX
XX 29-JUL-1997 (first entry)
XX
XX H. pylori cytoplasmic protein ORF 24824087.aa.
XX
XX Vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW cytoplasmic; ds.
XX
XX Helicobacter pylori.
OS
XX
XX Key Location/Qualifiers
XX CDS 1..531
XX FT /tag= a
XX FT /transl_except- (pos: 460..462, aa: Xaa)
XX FT /transl_except- (pos: 520..522, aa: Xaa)
XX FT /transl_except- (pos: 526..528, aa: Xaa)
XX FT /note= "Xaa = Unknown"

XX
XX WO9640893-A1.
XX
XX 19-DEC-1996.
XX
XX
XX 06-JUN-1996; 96WO-US09122.
XX
XX 01-APR-1996; 96US-0630405.
XX 07-JUN-1995; 95US-0487032.
XX
XX (ASTR) ASTRA AB.
XX
XX Berglindh OT, Smith D, Mellgaard BL;
PI
XX WPI: 1997-052306/05.
XX
XX P-PsDB; AAW20335.
XX
XX
XX Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
XX
XX Claim 9; Page -; 1481pp: English.
XX
XX The present sequence encodes a Helicobacter pylori cytoplasmic
CC protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the
CC bacterial DNA. The sequences were analysed for ORF of at least 180
CC nucleotides, and the predicted coding regions defined by computer
CC evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences
CC of interest, particular regions can be isolated from H. pylori by
CC PCR amplification for recombinant polypeptide production, e.g. in
CC E. coli hosts.
CC Note: This DNA sequence is not reproduced in the specification and
CC has been derived from the related specification, WO9719098.
XX
XX
XX Sequence 531 BP; 175 A; 88 C; 114 G; 151 T; 3 other;
SQ
XX
XX Query Match 1.8%; Score 18; DB 18; Length 531;
XX Best Local Similarity 100.0%; Pred. No. 30;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 572 aaagccgttcacccctt 589
DB 311 AAGCGCGCTTCACCCCTT 294
XX
XX
XX RESULT 18
XX AAT77453/C
XX ID AAT77453 standard; DNA; 531 BP.
XX
XX AAT77453;
XX
XX 11-AUG-1997 (first entry)
XX
XX H. pylori cytoplasmic protein ORF 24824087.aa.
XX
XX Chronic gastritis; duodenal ulcer disease; activator;
KW inhibitor; bacterial life cycle; vaccine; immunisation; detection;
KW antisense; inhibition; cytoplasmic; Na+/H+ antiporter;
KW Escherichia coli; ds.
XX
XX Helicobacter pylori.
OS
XX
XX Key Location/Qualifiers
XX CDS 1..531

```
FT      /*lag- a
FT      /transl_except- (pos: 460..462, aa: Xaa)
FT      /transl_except- (pos: 520..522, aa: Xaa)
FT      /transl_except- (pos: 526..528, aa: Xaa)
FT      /note= "Xaa = Unknown"
XX
XX      MO9719098-A1.
XX
XX      29-MAY-1997.
XX
XX      15-NOV-1996; 96WO-US18542.
XX
XX      17-NOV-1995; 95US-0561469.
XX
XX      (ASTR ) ASTRA AB.
XX
XX      Smith DH;
XX
XX      WPI: 1997-298052/27.
XX      P-PSDB; AAM24635.
XX
XX      Helicobacter pylori nucleic acid sequences and related proteins -
XX      used for diagnostics and therapeutics
XX
XX      Claim 1; Page 100; 235pp; English.
XX
XX      The present sequence encodes a Helicobacter pylori cytoplasmic
XX      protein, which was found to be homologous to Escherichia coli
XX      Na+/H+ antiporter protein following BLAST protein analysis.
XX      H. pylori has been strongly linked to chronic gastritis and
XX      duodenal ulcer disease. The nucleic acid sequences of the invention
XX      are used to evaluate compounds, especially activators or inhibitors
XX      of bacterial life cycle, for the ability to bind an H. pylori
XX      nucleic acid sequence. The nucleic acid sequences, and
XX      corresponding proteins, are also useful for generating vaccines for
XX      immunising subjects against H. pylori or for use in detecting the
XX      presence of Helicobacter species in a sample. Antisense nucleic
XX      acid sequences of these sequences are used to inhibit expression of
XX      a gene from Helicobacter species. H. pylori whole genomic DNA was
XX      isolated and nebulised to a median size of 2000 bp. Purified DNA
XX      fragments were blunt-ended and ligated to unique BstXI-linker
XX      adapters in 100-1000 fold molar excess. These linkers are
XX      complementary to the BstXI-cut PMPX vectors, while the overhang is
XX      not self-complementary. Therefore the linkers will not
XX      concatemrise nor will the cut vector re-ligate itself easily. The
XX      linker-adaptor inserts were ligated to each of the 20 PMPX vectors
XX      to construct a series of shotgun subclone libraries. The purified
XX      DNA samples were then sequenced.
XX      CC Note: The ORF/protein reference number for this sequence was
XX      obtained from the related specification, WO9640893.
XX
XX      SQ      Sequence 531 BP; 175 A; 88 C; 114 G; 151 T; 3 other;
XX
XX      Query Match      1.8%; Score 18; DB 18; Length 531;
XX      Best Local Similarity 100.0%; Pred. No. 30;
XX      Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      Oy      572 aaagcgcgtcacccctt 589
XX      |||||||
XX      Db      311 AAAGCGCTCACCCCTT 294
XX
XX      RESULT 19
XX      AAC44859
XX      ID      AAC44859 standard; DNA; 791 BP.
XX
XX      AC      AAC44859;
XX
XX      DT      18-OCT-2000 (first entry)
XX
XX      DE      Arabidopsis thaliana DNA fragment SEQ ID NO: 44401.
XX
```

```
KM      Hybridisation assay; genetic mapping; gene expression control;
KM      protein identification; signal transduction pathway;
KM      metabolic pathway; promoter; termination sequence; ss.
XX
XX      Arabidopsis thaliana.
XX      OS
XX      EP1033405-A2.
XX      PN
XX      06-SEP-2000.
XX
XX      PD
XX
XX      XX
XX      25-FEB-2000; 2000EP-0301439.
XX
XX      PF
XX      25-FEB-1999; 99US-0121825.
XX      05-MAR-1999; 99US-0123180.
XX      09-MAR-1999; 99US-0123548.
XX      23-MAR-1999; 99US-0125788.
XX      25-MAR-1999; 99US-0126264.
XX      29-MAR-1999; 99US-0126785.
XX      01-APR-1999; 99US-0127462.
XX      06-APR-1999; 99US-0128234.
XX      08-APR-1999; 99US-0128714.
XX      16-APR-1999; 99US-0129845.
XX      19-APR-1999; 99US-0130077.
XX      21-APR-1999; 99US-0130510.
XX      23-APR-1999; 99US-0130891.
XX      28-APR-1999; 99US-0131449.
XX      30-APR-1999; 99US-0132048.
XX      30-APR-1999; 99US-0132407.
XX      04-MAY-1999; 99US-0132485.
XX      05-MAY-1999; 99US-0132486.
XX      06-MAY-1999; 99US-0132487.
XX      07-MAY-1999; 99US-0132863.
XX      11-MAY-1999; 99US-0134256.
XX      14-MAY-1999; 99US-0134218.
XX      14-MAY-1999; 99US-0134219.
XX      14-MAY-1999; 99US-0134221.
XX      14-MAY-1999; 99US-0134370.
XX      18-MAY-1999; 99US-0134768.
XX      19-MAY-1999; 99US-0134941.
XX      20-MAY-1999; 99US-0135124.
XX      21-MAY-1999; 99US-0135353.
XX      24-MAY-1999; 99US-0135629.
XX      25-MAY-1999; 99US-0136021.
XX      27-MAY-1999; 99US-0136392.
XX      28-MAY-1999; 99US-0136782.
XX      01-JUN-1999; 99US-0137222.
XX      03-JUN-1999; 99US-0137528.
XX      04-JUN-1999; 99US-0137502.
XX      07-JUN-1999; 99US-0137724.
XX      08-JUN-1999; 99US-0138094.
XX      10-JUN-1999; 99US-0138540.
XX      10-JUN-1999; 99US-0138847.
XX      14-JUN-1999; 99US-0139119.
XX      16-JUN-1999; 99US-0139453.
XX      16-JUN-1999; 99US-0139453.
XX      17-JUN-1999; 99US-0139454.
XX      18-JUN-1999; 99US-0139454.
XX      18-JUN-1999; 99US-0139455.
XX      18-JUN-1999; 99US-0139456.
XX      18-JUN-1999; 99US-0139457.
XX      18-JUN-1999; 99US-0139458.
XX      18-JUN-1999; 99US-0139459.
XX      18-JUN-1999; 99US-0139460.
XX      18-JUN-1999; 99US-0139461.
XX      18-JUN-1999; 99US-0139462.
XX      18-JUN-1999; 99US-0139463.
XX      18-JUN-1999; 99US-0139750.
XX      18-JUN-1999; 99US-0139763.
XX      21-JUN-1999; 99US-0139817.
XX      22-JUN-1999; 99US-0139899.
XX      23-JUN-1999; 99US-0140353.
XX
```

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0148368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156456.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.8%; Score 18; DB 21; Length 791;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 929 ggaagatggttgaaga 946
|||||
Db 508 ggaagatggttgaaga 525

RESULT 20
AAAT67992/C
ID AAT67992 standard; DNA; 1593 BP.

XX AAT67992;

DT 16-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein ORF 06cp30603orf11.

XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacterium; life cycle; activator;

KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;

KW diagnosis; ds.

XX Helicobacter pylori.
OS
XX

```
FH Key Location/Qualifiers
FT CDS 1..1593
FT /tag= a
FT /note= "no stop codon given"
XX
XX WO9640893-A1.
XX
XX 19-DEC-1996.
XX
XX 06-JUN-1996; 96WO-US09122.
XX
XX 01-APR-1996; 96US-0630405.
XX 07-JUN-1995; 95US-0487032.
XX
XX (ASTR ) ASTRA AB.
XX
XX Berglindh OT, Smith D, Mellgaard BL;
XX
XX WPI: 1997-052306/05.
XX P-PSDB: AAM20739.
XX
XX Helicobacter pylori nucleic acid sequences and related
XX polypeptide(s) - useful for vaccines to treat or prevent H. pylori
XX infection, and to detect Helicobacter
XX
XX Claim 9; Page 833; 1481pp; English.
XX
XX The present sequence encodes a H. pylori cytoplasmic protein.
XX The protein may be used in a vaccine to prevent or treat H. pylori
XX infection or to identify H. pylori polypeptide binding compounds.
XX Useful as potential H. pylori life cycle activators or inhibitors.
XX The genomic sequence of H. pylori (ATCC 55679) was determined from
XX overlapping contigs generated by mechanically shearing the bacterial
XX DNA. The sequences were analysed for ORF of at least 180 nucleotides,
XX and the predicted coding regions defined by computer evaluation. To
XX identify likely H. pylori antigens for vaccine development, the amino
XX acid sequences predicted from various ORF were analysed for significant
XX homology to other known or exported membrane proteins. Having identified
XX and determined the sequences of interest, particular regions can be
XX isolated from H. pylori by PCR amplification for recombinant polypeptide
XX production, e.g. in E. coli hosts.
XX
XX Sequence 1593 BP; 541 A; 288 C; 334 G; 430 T; 0 other;
XX
XX Query Match 1.8%; Score 18; DB 18; Length 1593;
XX Best Local Similarity 100.0%; Pred. No. 30;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 572 aaagcgcctcacccctt 589
XX 1376 AAGCGCTTCACCCCTT 1359
XX
XX RESULT 21
XX AAC77190
XX ID AAC77190 standard; cDNA; 1600 BP.
XX
XX AAC77190;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF2745 polynucleotide sequence SEQ ID NO:5489.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnery; antiparietal; antiparkinsonian; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antineumatic; antihypertoid;
XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
```

```
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI: 2000-602362/57.
XX P-PSDB: AAB42981.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 4663-4664; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antiparietal; antiparkinsonian; neurotropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
XX antihypertoid; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 1600 BP; 376 A; 435 C; 474 G; 315 T; 0 other;
XX
XX Query Match 1.8%; Score 18; DB 21; Length 1600;
XX Best Local Similarity 100.0%; Pred. No. 30;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 464 gtcttggaaccctgct 481
XX 1360 gtcttggaaccctgct 1377
XX
XX RESULT 22
XX AAX08683/C
XX ID AAX08683 standard; cDNA; 1626 BP.
XX
XX AAX08683;
XX
XX 27-SEP-1999 (first entry)
```



```
XX Boehnisch B, Gallert K, Huels C, Muellner S;
PI WPI; 1999-527373/44.
XX
XX New insect cell vector containing the sequence encoding a
PT DEAD-superfamily protein, particularly a nucleic acid helicase, used
PT e.g. for identifying potential pharmaceuticals
XX
PS Disclosure: Page 38-40; 43pp; German.
XX
XX This invention describes the construction of a novel insect cell vector
CC (A) which contains a nucleic acid (I) that codes for a protein (II) of
CC the DEAD-superfamily (A), and recombinant insect viruses derived from
CC them, are used to express recombinant (II), particularly RNA and DNA
CC helicases. (II) are potentially useful for: (a) production of cell lines
CC for research into cancer, inflammation and apoptosis, or for clarifying
CC the mechanism of action of drugs, and (b) to identify pharmaceutical
CC activity in known compounds, e.g. anticancer and antiviral activities.
CC (II), which are difficult to express in bacteria and yeast, are expressed
CC at high level in insect cells, e.g. 300-400 mg per 109 cells. This
CC sequence encodes a human RNA-helicase p135 protein which is used in the
CC description of the method of the invention.
XX
SQ Sequence 4120 BP; 1099 A; 1021 C; 1236 G; 764 T; 0 other;

Query Match      1.8%; Score 18; DB 20; Length 4120;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 agagagagtgtagcagag 166
   |||||||
DB 809 AGGAGGAGTGGCAGAG 792

RESULT 25
AAA60239/c
ID AAA60239 standard; DNA; 17 BP.
XX
XX AAA60239;
AC
XX
XX 07-DEC-2000 (first entry)
DT
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 60.
XX
XX Human; mouse; prostate cancer predisposing gene; HPC2;
KM human chromosome 17p; gene therapy; peptide therapy; drug design;
KM PCR primer; sequencing primer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200027864-A1.
PN
XX
XX 18-MAY-2000.
PD
XX
XX 05-NOV-1999; 99WO-US26055.
PE
XX
XX 06-NOV-1998; 98US-0107468.
PR
XX
XX (MYRI-) MYRIAD GENETICS INC.
PA
XX
XX Tavligian SV, Teng DHF, Simard J, Rommens JM;
PI
XX
XX WPI; 2000-376481/32.
DR
XX
XX Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
XX Example 3; Page 56; 157pp; English.
PS
XX
XX The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
```

```
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 17 BP; 5 A; 2 C; 7 G; 3 T; 0 other;

Query Match      1.7%; Score 17; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 tectgccaatgccttc 693
   |||||||
DB 17 TCCTGCCAATGCTTC 1

RESULT 26
AAA60241/c
ID AAA60241 standard; DNA; 17 BP.
XX
XX AAA60241;
AC
XX
XX 07-DEC-2000 (first entry)
DT
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 62.
XX
XX Human; mouse; prostate cancer predisposing gene; HPC2;
KM human chromosome 17p; gene therapy; peptide therapy; drug design;
KM PCR primer; sequencing primer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200027864-A1.
PN
XX
XX 18-MAY-2000.
PD
XX
XX 05-NOV-1999; 99WO-US26055.
PE
XX
XX 06-NOV-1998; 98US-0107468.
PR
XX
XX (MYRI-) MYRIAD GENETICS INC.
PA
XX
XX Tavligian SV, Teng DHF, Simard J, Rommens JM;
PI
XX
XX WPI; 2000-376481/32.
DR
XX
XX Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
XX Example 3; Page 56; 157pp; English.
PS
XX
XX The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
XX Sequence 17 BP; 3 A; 5 C; 6 G; 3 T; 0 other;

Query Match      1.7%; Score 17; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 824 gcaacctcgtcgtga 840
   |||||||
DB 17 GCACACCTCTGCTGCA 1
```

```
RESULT 27
AAA60242
ID AAA60242 standard; DNA; 17 BP.
XX
AC AAA60242;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 63.
XX
KM Human; mouse; prostate cancer predisposing gene; HPC2;
KM human chromosome 17p; gene therapy; peptide therapy; drug design;
KM PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PE 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavrigian SV, Teng DHF, Simard J, Rommens JM;
DR WPI; 2000-376481/32.
XX
PT Human prostate cancer (HPC2) nucleic acids, polypeptides, and
XX antibodies; useful for treatment and diagnosis of prostate cancer -
XX
PS Example 3; Page 56; 157pp; English.
XX
CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 17 BP; 1 A; 5 C; 7 G; 4 T; 0 other;

Query Match 1.7%; Score 17; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 870 tgcgagctctgtcgcg 886
DB 1 tgcgagctctgtcgcg 17

RESULT 28
AAA32079/C
ID AAA32079 standard; DNA; 256 BP.
XX
AC AAA32079;
XX
DT 05-JUL-2000 (first entry)
XX
DE Plant microsatellite marker #1040.
XX
KM Plant microsatellite sequence; core repeat sequence; detection; probe;
KM DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KM variety identification; genetic variability evaluation; primer; ss.
XX
OS Pinus radiata.
```

```
XX
PN WO9967421-A1.
XX
PD 29-DEC-1999.
XX
PE 25-JUN-1999; 99WO-NZ00092.
XX
PR 25-JUN-1998; 98US-0105307.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Havukkala JU, Bloksberg LN, Glenn M;
DR WPI; 2000-116958/10.
XX
PT New plant microsatellite markers and associated flanking species for
XX the detection of polymorphic genetic markers -
XX
PS Claim 1; Page 382; 392pp; English.
XX
CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences AAA32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
SQ Sequence 256 BP; 89 A; 41 C; 35 G; 91 T; 0 other;

Query Match 1.7%; Score 17; DB 21; Length 256;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 833 tgcctggaagtgctct 849
DB 256 TGCCTGGAAGTGCTCT 240

RESULT 29
AAC32768/C
ID AAC32768 standard; DNA; 292 BP.
XX
AC AAC32768;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 560.
XX
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
```

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 9905-0158369.
 PR 13-OCT-1999; 9905-0159293.
 PR 13-OCT-1999; 9905-0159294.
 PR 13-OCT-1999; 9905-0159295.
 PR 14-OCT-1999; 9905-0159329.
 PR 14-OCT-1999; 9905-0159330.
 PR 14-OCT-1999; 9905-0159331.
 PR 14-OCT-1999; 9905-0159637.
 PR 14-OCT-1999; 9905-0159638.
 PR 18-OCT-1999; 9905-0159584.
 PR 21-OCT-1999; 9905-0160741.
 PR 21-OCT-1999; 9905-0160767.
 PR 21-OCT-1999; 9905-0160768.
 PR 21-OCT-1999; 9905-0160770.
 PR 21-OCT-1999; 9905-0160814.
 PR 21-OCT-1999; 9905-0160815.
 PR 22-OCT-1999; 9905-0160980.
 PR 22-OCT-1999; 9905-0160981.
 PR 22-OCT-1999; 9905-0160989.
 PR 25-OCT-1999; 9905-0161404.
 PR 25-OCT-1999; 9905-0161405.
 PR 25-OCT-1999; 9905-0161406.
 PR 26-OCT-1999; 9905-0161359.
 PR 26-OCT-1999; 9905-0161360.
 PR 26-OCT-1999; 9905-0161361.
 PR 28-OCT-1999; 9905-0161920.
 PR 28-OCT-1999; 9905-0161922.
 PR 28-OCT-1999; 9905-0161993.
 PR 29-OCT-1999; 9905-0162142.

Query Match
 Best Local Similarity 1.7%; Score 17; DB 21; Length 292;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 623 gctcagcagtagtaccaca 639
 DB 242 GCTCCAGCAGTACCACA 226

RESULT 30
 AAA32050/C
 ID AAA32050 standard; DNA; 300 BP.

AC AAA32050;
 XX
 DT 05-JUL-2000 (first entry)

DE Plant microsatellite marker #1011.

XX Plant microsatellite sequence; core repeat sequence; detection; probe;
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KW variety identification; genetic variability evaluation; primer; ss.

OS Pinus radiata.

PN W09967421-A1.

PD 29-DEC-1999.

PF 25-JUN-1999; 99WO-NZ00092.

PR 25-JUN-1998; 98US-0105307.

PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Hayukkala IJ, Bloksberg IN, Glenn M;

DR MPI, 2000-116958/10.

PT New plant microsatellite markers and associated flanking species for
 XX the detection of polymorphic genetic markers -

PS Claim 1; Page 373; 392pp; English.
 XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.

Query Match
 Best Local Similarity 1.7%; Score 17; DB 21; Length 300;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 833 tgctcggaagtgtct 849
 DB 292 TGCTCGAAGTGTCT 276

RESULT 31
 AA242643/C
 ID AA242643 standard; CDNA; 330 BP.

AC AA242643;
 XX
 DT 01-FEB-2000 (first entry)

DE Human 5' EST isolated from a CDNA library SEQ ID NO:402.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification; ss.

OS Homo sapiens.

PN W09953051-A2.

PD 21-OCT-1999.

PF 09-APR-1999; 99WO-IB00712.

PR 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

PA (GEST) GENSET.

PI Dunas Milne Edwards J, Duclert A, Giordano J;

DR MPI, 2000-038446/03.

DR P-PSDB; AAY65029.

PT Novel secreted protein 5' expressed sequence tag sequences used in
 XX diagnostic, forensic, gene therapy, and chromosome mapping procedures

PS Claim 1; Page 382; 837pp; English.

XX AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAY64651 to
 CC AAY65438 represent the EST-related proteins corresponding to AA242265 to
 CC AA243052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for

CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be
 CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value, and the identification of new secreted proteins is
 CC valuable. AA242249 to AA42264 and AA64644 to AA64650 represent
 CC sequences used in the exemplification of the present invention.

SO Sequence 330 BP; 76 A; 102 C; 86 G; 65 T; 1 other;

Query Match 1.7%; Score 17; DB 21; Length 330;

Best Local Similarity 100.0%; Pred. No. 96;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 acctcatccaccgcgac 46
 |||||||
 DB 48 ACCTCATCCACCGGAC 32

RESULT 32

ID AAT24629 standard; cDNA to mRNA; 347 BP.

AC AAT24629;

DT 07-OCT-1996 (first entry)

DE Human gene signature HUMGS06689.

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KM human; cloning; mapping; non-biased library; diagnosis; detection;

KM cell typing; abnormal cell function; ss.

OS Homo sapiens.

XX WO9514772-A1.

XX 01-JUN-1995.

PE 11-NOV-1994; 94WO-JP01916.

PR 12-NOV-1993; 93JP-0355504.

PA (MATS/) MATSUBARA K.

PA (OKUBO/) OKUBO K.

PI Matsubara K, Okubo K;

DR MPI; 1995-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

XX tissues

PS Claim 1; Page 1655; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in AAT19001-T26837 and which is able to hybridize to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

SO Sequence 347 BP; 61 A; 102 C; 108 G; 75 T; 1 other;

Query Match 1.7%; Score 17; DB 16; Length 347;

Best Local Similarity 100.0%; Pred. No. 96;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 147 ccaggaggagtgccag 163
 |||||||
 DB 15 ccaggaggagtgccag 31

RESULT 33

ID AAC07468 standard; cDNA; 350 BP.

AC AAC07468;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 11543.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KM gene therapy; chromosome mapping; ss.

OS Homo sapiens.

XX EP1033401-A2.

PN 06-SEP-2000.

PD 21-FEB-2000; 2000EP-0200610.

PF 26-FEB-1999; 99US-0122487.

PR (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX MPI; 2000-500381/45.

DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 11543; 71pp + CD-ROM; English.

PS The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. No ORF has yet been conclusively

CC identified within the present sequence. The 5' ESTs were prepared from

CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

CC sequences usually correspond mainly to the 3' untranslated region (UTR)

CC of the mRNA because they are often obtained from oligo-dT primed cDNA

CC libraries. Such ESTs are not well suited for isolating cDNA sequences

CC derived from the 5' ends of mRNAs and even in those cases where longer

CC cDNA sequences have been obtained, the full 5' UTR is rarely included.

CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.

CC They are used to obtain upstream regulatory sequences and to design

CC expression and secretion vectors.

XX Sequence 350 BP; 96 A; 82 C; 78 G; 94 T; 0 other;

SO Query Match 1.7%; Score 17; DB 21; Length 350;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 158 tggcagagggatgccat 174
 |||
 Db 91 tggcagagggatgccat 107

RESULT 34

AAAF64559
 ID AAFF64559 standard; CDNA; 398 BP.

AAFF64559;

09-APR-2001 (first entry)

Novel human polynucleotide, SEQ ID NO: 315.

Human; cytostatic; gene therapy; colon cancer; prostate cancer;
 breast cancer; lung cancer; cancer detection; ss.

Homo sapiens.

MO200102568-A2.

11-JAN-2001.

30-JUN-2000; 2000MO-US18374.

02-JUL-1999; 99US-0142310.

02-JUL-1999; 99US-0142311.

(CHIR) CHIRON CORP.

(HYSE-) HYSEQ INC.

Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
 Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Dirmannac R;
 Cirenjakov R, Dirmannac S, Dickson M, Labat I, Leshkowitz D;
 Kita D, Garcia V, Jones LM, Strache-Crain B;

WPI; 2001-091805/10.

Library of polynucleotides for diagnosing a cancerous state of a
 mammalian cell and detecting cancer, particularly of the colon or
 prostate, comprises 3351 human polynucleotide sequences -

Claim 9; Page 590; 1046pp; English.

The present sequence is one of 3351 sequences in a library of human
 polynucleotides. The library is used to detect differentially expressed
 genes correlated with a cancerous state of a mammalian cell and can
 detect colon, prostate, breast and lung cancer. The library can be used
 to produce probes for detection of mRNA and to produce additional copies
 of the polynucleotides. The probes can be used for chromosome mapping of
 the polynucleotide and for detection of transcription levels. Ribozymes
 or antisense oligonucleotides can be generated. The polynucleotides and
 their gene products are used as genetic or biochemical markers (e.g. in
 blood or tissues) that will detect the earliest changes along the
 carcinogenesis pathway and/or monitor the efficacy of therapies and
 preventive interventions. The polynucleotides, polypeptides and
 antibodies against them can be used in pharmaceutical compositions to
 treat the cancers and proliferative disorders such as neoplasia,
 dysplasia and hyperplasia.

Sequence 398 BP; 65 A; 105 C; 153 G; 73 T; 2 other;

Query Match 1.7%; Score 17; DB 22; Length 398;

Best Local Similarity 100.0%; Pred. No. 96;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 462 gggtctctgggacacctg 478
 |||
 Db 130 gggtctctgggacacctg 146

RESULT 35
 AAA32069/C
 ID AAA32069 standard; DNA; 426 BP.

AAA32069;

05-JUL-2000 (first entry)

Plant microsatellite marker #1030.

Plant microsatellite sequence; core repeat sequence; detection; probe;
 DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 variety identification; genetic variability evaluation; primer; ss.

Pinus radiata.

WO9967421-A1.

29-DEC-1999.

25-JUN-1999; 99MO-NZ00092.

25-JUN-1999; 98US-0105307.

(GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.

(FLET-) FLETCHER CHALLENGE FORESTS LTD.

Havukkala IJ, Bloksberg LN, Glenn M;

WPI; 2000-116958/10.

New plant microsatellite markers and associated flanking species for
 the detection of polymorphic genetic markers -

Claim 1; Page 379; 392pp; English.

Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 and associated flanking species. The sequences comprise a central core
 repeat sequence, especially selected from the sequences AAA32094-A32096
 with left and right flanking sequences. The polynucleotide sequences
 can be used in the detection of DNA polymorphisms, in genome mapping,
 in physical mapping, in positional cloning of genes, in variety
 identification and in evaluation of genetic variability within and
 between plant tissues, populations, cultivars, species and species
 groups. They may also be used to design hybridization probes for
 oligonucleotide fingerprinting and library screening and to design
 primers for microsatellite-primed PCR. Microsatellite markers are
 useful to locate specific economically useful genes in plant genomes.

Sequence 426 BP; 124 A; 81 C; 71 G; 150 T; 0 other;

Query Match 1.7%; Score 17; DB 21; Length 426;

Best Local Similarity 100.0%; Pred. No. 96;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 833 tggctggaaagtggtct 849
 |||
 Db 257 TGGCTGGAAGTGCT 241

RESULT 36

AAA32052/C
 ID AAA32052 standard; DNA; 431 BP.

AAA32052;

05-JUL-2000 (first entry)

Plant microsatellite marker #1013.

KM Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KM variety identification; genetic variability evaluation; primer; ss.
OS Pinus radiata.
PN WO967421-A1.
XX 29-DEC-1999.
PD 25-JUN-1999; 99WO-N200092.
XX 25-JUN-1999; 98US-0105307.
PR 25-JUN-1998; 98US-0105307.
XX (GENE-) GENESTS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX Havukkala IJ, Bloksberg LN, Glenn M;
PI WPI: 2000-116958/10.
XX
DR New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -
XX
XX Claim 1: Page 374; 392pp: English.
PS
XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences AAA32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
SQ Sequence 431 BP; 129 A; 78 C; 71 G; 152 T; 1 other;

Query Match 1.7%; Score 17; DB 21; Length 431;
Best Local Similarity 100.0%; Pred No. 96;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 833 tgagctggaagtgtct 849
|||
Db 262 TGCGTGGAAAGTGTCT 246

RESULT 37
AAC46096/c
ID AAC46096 standard; DNA; 459 BP.
XX
AC AAC46096;
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 48895.
XX
XX
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PF
XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.7%; Score 17; DB 21; Length 459;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 623 gctccagcagttaccaca 639
Db 179 gctccagcagttaccaca 163

RESULT 38
AAC45096/c
ID AAC45096 standard; DNA; 523 BP.
XX
XX AAC45096;
AC
AC 18-OCT-2000 (first entry)
DT
DT
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 45285.
DE
DE Arabidopsis thaliana
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
OS
PN
PN EP1033405-A2.
PD
PD
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140033.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144431.
PR 19-JUL-1999; 99US-0144431.

PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154387.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157153.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.

```
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159384.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

```
Query Match 1.7%; Score 17; DB 21; Length 523;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 623 gctccagcagtlaccaca 639
DB 243 GCTCCAGCAGTACCACA 227
```

```
RESULT 39
AAV29360/C
ID AAV29360 standard; DNA; 559 BP.
XX
AC AAV29360;
XX
DT 31-JUL-1998 (first entry)
XX
DE Calcium ion channel alpha subunit exons 33, 34/intron partial sequence.
XX
KW Calcium ion channel alpha subunit; human; episodic ataxia type 2;
KW familial hemiplegic migraine; FHM; EA-2; treatment; diagnosis;
KW exon; Intron; ss.
XX
OS Homo sapiens.
XX
FH Key 1.156 Location/Qualifiers
FT intron /tag= a
FT /number= 32
FT /note= "partial sequence"
FT exon 157..222
FT /tag= b
FT /number= 33
FT intron 223..394
FT /tag= c
FT /number= 33
FT exon 395..509
FT /tag= d
FT /number= 34
FT intron 510..559
FT /tag= e
FT /number= 34
FT /note= "partial sequence"
XX
XX EP934561-A1.
```

```
XX 08-APR-1998.
PD
XX
XX 27-SEP-1996; 96EP-0202707.
PF
XX 27-SEP-1996; 96EP-0202707.
PR
XX (UYLE-) RIJCSUNIV LEIDEN.
XX
XX Ferrari MD, Frants RRE, Ophoff RA, Terwindt GM;
PI WPT; 1998-195461/18.
XX
XX New human nucleic acid associated with migraine and episodic ataxia
PT type 2 - useful for diagnosis and development of specific treatments
XX
XX Disclosure; Fig 1; 157pp; English.
XX
XX Sequences shown in AAV29330 to AAV29371 represent the 47 exons and
CC flanking intronic sequences containing the complete coding region of the
CC human calcium ion channel alpha 1 subunit gene and part of untranslated
CC sequences. The channel is related to familial hemiplegic migraine (FHM)
CC and/or episodic ataxia type 2 (EA-2) and is derived from, related to or
CC associated with a gene present in humans on chromosome 19p13.1-13.2. The
CC encoding gene can be used to localise or identify genes related to
CC episodic neurological disorders, specifically migraine, FHM or EA-2, but
CC also epilepsy. The isolated or a recombinant nucleic acid can also be
CC used to distinguish between alleles of the corresponding gene. Cells and
CC animals containing recombinant expression vectors comprising the nucleic
CC acid can be useful in study, development and treatment of migraine, FHM,
CC EA-2 and epilepsy. Proteins or peptides encoded by the nucleic acid and
CC natural or synthetic antibodies against the proteins can be used to
CC diagnose FHM, EA-2, migraine and other neurological conditions
CC associated with cation channel dysfunction.
XX
XX Sequence 559 BP; 119 A; 157 C; 160 G; 120 T; 3 other;
```

```
Query Match 1.7%; Score 17; DB 19; Length 559;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 aaaccagctcaactc 34
DB 339 AAACCAGCTCAACCTC 323

RESULT 40
AAf64735
ID AAF64735 standard; CDNA; 694 BP.
XX
AC AAF64735;
XX
DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, seq ID NO: 491.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
XX WO200102568-A2.
XX
XX 11-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18374.
XX
XX 02-JUL-1999; 99US-0142310.
XX
XX 02-JUL-1999; 99US-0142311.
XX
XX (CHIR ) CHIRON CORP.
XX
XX (HYSE-) HYSEQ INC.
PA
```

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkjenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LM, Strache-Crain B;
XX
DR WPI: 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9: Page 614-615; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 694 BP; 127 A; 176 C; 246 G; 140 T; 5 other;

Query Match
Best Local Similarity 1.7%; Score 17; DB 22; Length 694;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 gggctcctggcaccctg 478
DB 134 gggctcctggcaccctg 150

RESULT 41
AAZ24979/C
ID AAZ24979 standard; CDNA; 791 BP.
XX
AC AAZ24979;
XX
DT 05-JUL-1999 (first entry)
XX
DE Clone GJ156 encoding TRAIN-R secreted form C-terminus.
XX
XX TRAIN-R; receptor; human; tumour necrosis factor receptor;
KW agonist; antagonist; cancer; immunological disease; therapy;
KW cytostatic; ss.
XX
OS Homo sapiens.
XX
EH Key Location/Qualifiers
FT Intron 1..350
FT Exon /*tag= a
FT CDS 351..790
FT /*tag= b
FT /*tag= c
FT /*tag= d
FT 3'UTR /*tag= "TRAIN-R secreted form C-terminus"
XX
XX WO913078-A1.
XX
XX 18-MAR-1999.

XX
XX 11-SEP-1998; 98WO-US19030.
XX
XX 06-MAY-1998; 98US-0084422.
XX
XX 12-SEP-1997; 97US-0038631.
XX
XX
XX (BIOJ) BIOGEN INC.
XX
XX Hession C, Tschopp J;
XX
XX WPI: 1999-229238/19.
XX
XX P-PSDB; AAW98147.
XX
PT New cysteine-rich tumor necrosis factor receptor
XX
PS Claim 1: Page 28; 30pp; English.
XX
CC The present sequence includes an exon encoding the C-terminus (see
CC AAW98147) of a soluble form of a novel human cysteine-rich tumour
CC necrosis factor receptor family member termed TRAIN-R. It comprises
CC clone GJ156, obtained from a Clontech human adult lung cDNA library.
CC The encoded 30-amino acid C-terminal peptide is identical to amino
CC acids 121-149 of the composite TRAIN-R protein given in AAW98146 and
CC to amino acids 121-150 of the C-terminus of murine TRAIN-R short
CC form (secreted protein, see AAW98144). The soluble protein is
CC expected to inhibit signalling by the full-length TRAIN-R. Human
CC TRAIN-R is expressed at low levels in every tissue and cell line
CC tested thus far, with higher expression detected in heart, prostate,
CC ovary, testis, peripheral blood lymphocytes, thyroid and adrenal
CC gland. Cell death can be induced by administering an agent capable
CC of inhibiting the binding of TRAIN-R to its ligand. A claimed method
CC of treating, or reducing, the advancement, severity or effects of
CC an immunological disease in a mammal comprises administering a
CC pharmaceutical composition which comprises a TRAIN-R blocking agent,
CC e.g. soluble TRAIN-R. TRAIN-R can be fused to an immunoglobulin to
CC produce a fusion protein which may be targeted to various sites.
CC It can be used in binding assays, and to identify antagonists and
CC agonists. Anti-TRAIN-R antibodies can be used to reduce the
CC severity of an immune response or to treat cancer. TRAIN-R
CC blocking agents can also be used to reduce the severity or effects
CC of an immunological disease (all claimed).
XX
SQ Sequence 791 BP; 202 A; 189 C; 165 G; 235 T; 0 other;

Query Match
Best Local Similarity 1.7%; Score 17; DB 20; Length 791;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 942 gaagaggaaagcactgga 958
DB 203 GAAGAGGAAAGCAGCTGGA 187

RESULT 42
AAZ91823
ID AAZ91823 standard; DNA; 855 BP.
XX
AC AAZ91823;
XX
DT 02-JUN-2000 (first entry)
XX
DE Streptococcus pneumoniae DNA sequence ID33.
XX
XX Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW pneumococcal septicaemia; sinusitis; meningitis; therapy; ss.
XX
XX Streptococcus pneumoniae.
XX
XX
XX WO200006738-A2.
XX

10-FEB-2000 .
99WO-GB02452.
27-JUL-1999; ,
27-JUL-1998; 98GB-0016336.
19-MAR-1999; 99US-0125329.
(MCCR-) MICROBIAL TECHNIQS LTD.
Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
WPI: 2000-195301/17.
P-PSDB; AAW81727.
Streptococcal proteins and polynucleotides useful for diagnosis,
treatment and prophylaxis of bacterial infections
Claim 2: Page 47-48; 76pp; English.
This sequence encodes a Streptococcus pneumoniae protein of the
invention. The proteins (or their homologues, derivatives and/or
fragments) are useful as immunogens or antigens. Immunogenic or antigenic
compositions comprising the proteins are useful as vaccines and also in
diagnostic assays. The sequences are useful for the detection or
diagnosis of S. pneumoniae infection, by contacting a sample to be tested
with them. Agents capable of antagonising, inhibiting or interfering with
the function or expression of the protein or polypeptide are useful in
medical compositions in the treatment or prophylaxis of S. pneumoniae
infection. As the sequences can be used to treat S. pneumoniae infection,
they can be used to treat bacterial pneumonia, which has high rates in
young children, the elderly, and in patients with predisposing conditions
such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
or with immunosuppressive disorders, especially AIDS. They can also be
used to treat pneumococcal septicaemia, otitis media, sinusitis, and
meningitis.
Sequence 855 BP; 235 A; 173 C; 202 G; 245 T; 0 other;

PR 10-Oct-1997; 97US-0061765.
PR 10-Oct-1997; 97US-0948705.
PR 11-Oct-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Haare M, Obata Y, Old LJ;
PI Pfeundschn M, Sahn U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI: 1999-132448/11.
XX P-PSDB: AAY06998.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX
PS Claim 67; Page 780; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX Sequence 1022 BP; 230 A; 273 C; 302 G; 217 T; 0 other;

5Q Sequence 1022 BP; 230 A; 273 C; 302 G; 217 T; 0 other;

Query Match	1.7%	Score 17;	DB 21;	Length 855;
Best Local Similarity	100.0%	Pred. No. 95;		
Matches 17; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

Query Match	1.7%;	Score 17;	DB 20;	Length 1022;
Best Local Similarity	100.0%;	Pred. No. 95;		
Matches 17;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	759	gatttgaagagtttca	775
Db	139	gatttgaagagtttca	155

QY	463	ggtcctgggcacctgg	479
Db	203	ggtcctgggcacctgg	219

```

RESULT  43
AAK40199
ID      AAK40199 standard; DNA; 1022 BP.

```

```

RESULT  44
AAQ67866
ID  AAQ67866 standard; DNA; 1084 BP

```

AC AAX40199;

22-MAR-1995 (first entry)

DT 02-JUL-1999 (first entry)

H6/MAGE-1 expression cassette from pMAW037.

MAGE-4 encoding gene

Polymerase chain reaction; primer; amplify; NYVAC; ALVAC; recombinant;

KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer; ss.

primary melanoma tumour cell; melanoma-derived cell line; tumour; poxvirus; antigenic response; immunological response; pathogen; ss
Synthetic.

OS Homo sapiens.

Synthetic

PN W09904265-A2.

Key	Location/Qualifiers
misc_feature	1..51

PD 28-JAN-1999.

promoter	5'	3'	Flanking sequence"
	52..178		

PF 15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

```

/note= "Vaccinia H6 promoter"
179..1009
CDS

```

PR 10-OCT-1997; 97US-0061599.

```

/.../mag = C
FT /product= MAGE-1

```

```

FT misc_feature 1010..1084
FT /tag= d
FT /note= "Flanking sequence"
XX
XX WO9416716-A.
XX
XX 04-AUG-1994.
XX
XX 21-JAN-1994; 94WO-US00888.
XX
XX 21-JAN-1993; 93US-0007115.
XX
XX 19-JAN-1994; 94US-0184009.
XX
XX (VIRO-) VIROGENETICS CORP.
XX
XX Cox WI, Paoletti E, Tartaglia J;
XX
XX WPI; 1994-263767/32.
XX
XX Attenuated recombinant virus used for cancer therapy - comprises
XX DNA encoding cytokine and/or tumour associated antigen
XX
XX Example 16; Fig 20; 232pp; English.
XX
XX The sequences given in AAQ67865-66 represent expression cassettes
XX containing the vaccinia H6 promoter and the human MAGF-1 gene which
XX encodes human melanoma-associated antigen M22-E, in vcr235 and pMAW037,
XX respectively. These sequences were used in the construction of NYVAC-
XX and ALVAC-based recombinant viruses containing the MAGF-1 gene. MAGF-1
XX is expressed in primary melanoma tumour cells, melanoma-derived cell
XX lines and certain tumours of non-melanoma origins but not in normal
XX cells except in testis. A first PCR fragment containing the last 18 bp
XX and the initial 24 nucleotides of the MAGF-1 gene was generated and
XX ligated to a second PCR fragment amplified from plasmid pTZ18MAGE1
XX which contains the initial 546 bp of the MAGF-1 coding sequence. The
XX terminal sequence of MAGF-1 was amplified and a fusion product was
XX generated containing the H6 promoter and the full length MAGF-1
XX sequence. This construct may be introduced in to the poxvirus derived
XX plasmid, ALVAC and NYVAC. The resulting viruses may be used in a
XX composition for inducing an antigenic or immunological response, ie. for
XX immunisation against pathogens.
XX
XX Sequence 1084 BP; 266 A; 256 C; 280 G; 282 T; 0 other;
XX
XX
XX Query Match
XX Best Local Similarity 1.7%; Score 17; DB 15; Length 1084;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 463 ggtcctgggacacctgg 479
XX ||||||||||||||||
XX Db 292 ggtcctgggacacctgg 308
XX
XX
XX RESULT 45
XX AA208442
XX ID AA208442 standard; DNA; 1084 BP.
XX
XX AC AA208442;
XX
XX DT 19-OCT-1999 (first entry)
XX
XX DE H6/MAGF-1 expression cassette and flanking regions from pMAW037.
XX
XX Attenuated recombinant virus; cytokine; tumour associated antigen;
XX NYVAC recombinant virus; ALVAC recombinant virus; gene therapy; rabies;
XX cancer; tumour necrosis factor; nuclear phosphoprotein; p53; IL-2; GM-CSF;
XX interleukin; interferon; IFN-gamma; IL-4; melanoma associated antigen;
XX carcinoembryonic antigen; immunisation; poxvirus; influenza;
XX immunological response; immunotherapy; vaccine; Newcastle Disease; ss.
XX
XX Synthetic.
XX OS Homo sapiens.

```

```

OS Vaccinia virus.
XX
XX US5942235-A.
XX
XX PD 24-AUG-1999.
XX
XX PF 02-JUN-1995; 95US-0458356.
XX
XX 02-JUN-1995; 95US-0458356.
XX
XX 02-JUN-1995; 95US-0458356.
XX
XX 02-DEC-1981; 81US-034456.
XX
XX 08-DEC-1982; 82US-0446824.
XX
XX 19-JUN-1984; 84US-0622135.
XX
XX 27-AUG-1987; 87US-0090209.
XX
XX 28-AUG-1987; 87US-0090711.
XX
XX 20-OCT-1987; 87US-0110335.
XX
XX 25-APR-1988; 88US-0186054.
XX
XX 23-AUG-1988; 88US-0234390.
XX
XX 08-MAR-1989; 89US-0320471.
XX
XX 14-FEB-1990; 90US-0478179.
XX
XX 14-JUN-1990; 90US-0537882.
XX
XX 14-JUN-1990; 90US-0537890.
XX
XX 07-JAN-1991; 91US-0638080.
XX
XX 07-MAR-1991; 91US-0666056.
XX
XX 11-JUN-1991; 91US-0713967.
XX
XX 16-DEC-1991; 91US-0805567.
XX
XX 03-MAR-1992; 92US-0847977.
XX
XX 06-MAR-1992; 92US-0847951.
XX
XX 04-MAY-1992; 92US-0881995.
XX
XX 22-JUL-1992; 92US-0918278.
XX
XX 20-JAN-1993; 93US-0007115.
XX
XX 19-JAN-1994; 94US-0184009.
XX
XX 14-APR-1994; 94US-0228926.
XX
XX 13-SEP-1994; 94US-0306259.
XX
XX
XX (HEAL-) HEALTH RES INC.
XX
XX Paoletti E;
XX
XX WPI; 1999-493494/41.
XX
XX
XX Recombinant poxviruses comprising exogenous DNA encoding antigenic
XX determinants useful in immunotherapy to immunize against cancers and
XX other diseases such as influenza, Newcastle Disease and rabies
XX
XX Example 16; Fig 20; 163pp; English.
XX
XX The present invention describes a recombinant poxvirus (I), comprising
XX exogenous DNA encoding an antigenic determinant of a pathogen which is
XX then expressed in vivo in infected host cells after administration to a
XX patient and therefore induces an immunological response. (I) may be used
XX to vaccinate patients against a wide range of diseases and disorders
XX depending on the type of antigen encoded by the exogenous DNA. (I) may
XX be used to vaccinate against diseases such as rabies, influenza and
XX Newcastle Disease. It is particularly useful for immunising against
XX cancers. The poxvirus (I) also provides a means of manipulating
XX lymphocytes and tumour cells for use in cell-based immunotherapeutic
XX modalities for cancer. (I) also have enhanced safety compared to
XX unattenuated viruses (attenuation reduces the virulence of the viruses)
XX and known recombinant poxvirus vaccines. This increased level of safety
XX reduces the possibility of a 'runaway' infection in the host and reduces
XX the chance of transmission from vaccinated to unvaccinated individuals
XX and contamination of the environment. The present sequence represents a
XX H6/MAGF-1 expression cassette and flanking regions from pMAW037
XX used in the exemplification of the present invention.
XX
XX Sequence 1084 BP; 266 A; 256 C; 280 G; 282 T; 0 other;
XX
XX
XX Query Match
XX Best Local Similarity 1.7%; Score 17; DB 20; Length 1084;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 463 ggtcctgggacacctgg 479

```

Db 292 ggtcctggtgacccctgg 308

Search completed: October 28, 2001, 22:35:36
Job time: 2629 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 21:50:02 ; Search time 51.72 seconds

(without alignments)
3663.968 Million cell updates/sec

Title: SEQ1INS_COPY_1140_2140

Perfect score: 1001
1 tcgcagccacaagattcaaa.....aagccatcagcgtgsgatg 1001

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size : 8

Total number of hits satisfying chosen parameters: 255877

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- Issued_Patents_NA:*
- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
 - 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
 - 3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
 - 4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
 - 5: /cgn2_6/prodata/2/ina/PCrUS_COMB.seq:*
 - 6: /cgn2_6/prodata/2/ina/backfilesl.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	1.8	2186	2 US-08-878-546-9	Sequence 9, Appl
2	17	1.7	289	2 US-08-967-101-23	Sequence 23, Appl
3	17	1.7	289	2 US-08-592-541-23	Sequence 23, Appl
4	17	1.7	289	3 US-09-124-698-23	Sequence 23, Appl
5	17	1.7	289	4 US-09-127-880-23	Sequence 23, Appl
6	17	1.7	289	4 US-08-496-841C-23	Sequence 23, Appl
7	17	1.7	1084	2 US-08-184-009-110	Sequence 110, App
8	17	1.7	1084	2 US-08-458-356-110	Sequence 110, App
9	17	1.7	1094	2 US-08-184-009-109	Sequence 109, App
10	17	1.7	1094	2 US-08-458-356-109	Sequence 109, App
11	17	1.7	1330	4 US-09-118-442-29	Sequence 29, Appl
12	17	1.7	1691	2 US-08-993-118-8	Sequence 8, Appl
13	17	1.7	1691	3 US-08-845-528C-8	Sequence 8, Appl
14	17	1.7	2419	1 US-07-807-043B-7	Sequence 7, Appl
15	17	1.7	2419	1 US-08-299-849B-7	Sequence 7, Appl
16	17	1.7	2419	2 US-08-142-368A-7	Sequence 7, Appl
17	17	1.7	2419	3 US-08-967-727-7	Sequence 7, Appl
18	17	1.7	2419	4 US-08-037-230D-7	Sequence 7, Appl
19	17	1.7	2420	1 US-08-465-167A-23	Sequence 23, Appl
20	17	1.7	4488	1 US-08-441-430-1	Sequence 1, Appl
21	17	1.7	5674	1 US-07-807-043B-8	Sequence 8, Appl
22	17	1.7	5674	1 US-08-190-411A-1	Sequence 1, Appl
23	17	1.7	5674	1 US-08-289-849B-8	Sequence 8, Appl
24	17	1.7	5674	2 US-08-560-024-1	Sequence 1, Appl
25	17	1.7	5674	2 US-08-142-368A-8	Sequence 8, Appl
26	17	1.7	5674	3 US-08-967-727-8	Sequence 8, Appl
27	17	1.7	5674	4 US-08-037-230D-8	Sequence 8, Appl

28	17	1.7	11236	1 US-07-853-913-1	Sequence 1, Appl
29	17	1.7	68750	3 US-09-335-409-1	Sequence 1, Appl
30	16	1.6	76	1 US-07-753-110B-12	Sequence 12, Appl
31	16	1.6	76	1 US-08-503-730-16	Sequence 16, Appl
32	16	1.6	76	2 US-08-507-634-13	Sequence 13, Appl
33	16	1.6	235	4 US-08-905-223-207	Sequence 207, Appl
34	16	1.6	252	2 US-08-630-822A-97	Sequence 97, Appl
35	16	1.6	252	2 US-09-005-069-97	Sequence 97, Appl
36	16	1.6	252	3 US-08-906-616-104	Sequence 104, App
37	16	1.6	252	3 US-08-906-616-104	Sequence 104, App
38	16	1.6	252	3 US-08-817-795-104	Sequence 104, App
39	16	1.6	252	3 US-08-639-075A-104	Sequence 104, App
40	16	1.6	252	4 US-09-012-431-104	Sequence 104, App
41	16	1.6	252	4 US-09-012-692-104	Sequence 104, App
42	16	1.6	252	4 US-08-906-613-104	Sequence 104, App
43	16	1.6	252	5 PCT-US95-14442A-104	Sequence 104, App
44	16	1.6	291	4 US-09-060-756-254	Sequence 254, App
45	16	1.6	489	1 US-08-334-254-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-08-878-546-9
; Sequence 9, Application US/08878546
; Patent No. 5952463
; GENERAL INFORMATION:
; APPLICANT: SHIBANO, YUJI
; APPLICANT: KIKUCHI, NORIHISA
; APPLICANT: ODA, KOHEI
; TITLE OF INVENTION: NOVEL PROTEINASE INHIBITOR AND
; TITLE OF INVENTION: GENE ENCODING THE INHIBITOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: STEINBERG, RASKIN & DAVIDSON P.C.
; STREET: 1140 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,546
; FILING DATE: 19-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 158677/1996
; FILING DATE: 19-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 224104/1996
; FILING DATE: 26-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 48101/1997
; FILING DATE: 03-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: DAVIDSON, CLIFFORD M.
; REGISTRATION NUMBER: 32,728
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)-768-3800
; TELEFAX: (212)-382-2124
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: STREPTOMYCES PLATENSIS
 ; STRAIN: Q268
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1477..1911
 ; US-08-878-546-9

Query Match 1.8%; Score 18; DB 2; Length 2186;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 ggtcctggcaccctggc 480
 DB 1518 GGCTCTGGCACCCTGGC 1535

RESULT 2
 US-08-967-101-23/c
 ; Sequence 23, Application US/08967101
 ; Patent No. 5840540
 ; GENERAL INFORMATION:
 ; APPLICANT: ST. GEORGE-HYSLOP, PETER H
 ; APPLICANT: ROMMENS, JOHANNA M
 ; APPLICANT: FRASER, PAUL E
 ; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
 ; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
 ; NUMBER OF SEQUENCES: 183
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
 ; STREET: High Street Tower - 125 High Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/967,101
 ; FILING DATE: 10-NOV-1997
 ; CLASSIFICATION: 435
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: 08/592,541
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pletcher, Edmund R.
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 248-7000
 ; TELEFAX: (617) 248-7100
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 289 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-967-101-23

Query Match 1.7%; Score 17; DB 2; Length 289;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 aagaggaagcagtgaa 959
 DB 267 AAGAGGAAGCAGTGAA 251

RESULT 3
 US-08-592-541-23/c
 ; Sequence 23, Application US/08592541
 ; Patent No. 5986054
 ; GENERAL INFORMATION:
 ; APPLICANT: ST. GEORGE-HYSLOP, PETER H
 ; APPLICANT: ROMMENS, JOHANNA M
 ; APPLICANT: FRASER, PAUL E
 ; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
 ; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
 ; NUMBER OF SEQUENCES: 183
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
 ; STREET: High Street Tower - 125 High Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/592,541
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pletcher, Edmund R.
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 248-7000
 ; TELEFAX: (617) 248-7100
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 289 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-592-541-23

Query Match 1.7%; Score 17; DB 2; Length 289;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 aagaggaagcagtgaa 959
 DB 267 AAGAGGAAGCAGTGAA 251

RESULT 4
 US-09-124-698-23/c
 ; Sequence 23, Application US/09124698
 ; Patent No. 6117978
 ; GENERAL INFORMATION:
 ; APPLICANT: ST. GEORGE-HYSLOP, PETER H
 ; APPLICANT: ROMMENS, JOHANNA M
 ; APPLICANT: FRASER, PAUL E
 ; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
 ; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
 ; NUMBER OF SEQUENCES: 183
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
 ; STREET: High Street Tower - 125 High Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible


```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/124,698
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ. ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-124-698-23

```

```

Query Match      1.7%; Score 17; DB 3; Length 289;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 943 aagagaagcagtgaa 959
Db 267 AAGAGAACGACGTGAA 251

```

```

RESULT 5
US-09-127-480-23/c
; Sequence 23, Application US/09127480
; Patent No. 6194153
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/127,480
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ. ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-127-480-23

```

```

Query Match      1.7%; Score 17; DB 4; Length 289;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 943 aagagaagcagtgaa 959
Db 267 AAGAGAACGACGTGAA 251

```

```

RESULT 6
US-08-496-841C-23/c
; Sequence 23, Application US/08496841C
; Patent No. 6210919
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dardy & Dardy, PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/496,841C
; FILING DATE: 28-Jun-1995
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul F. Fehlner, Ph.D.
; REGISTRATION NUMBER: 35,135
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ. ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-08-496-841C-23

```

```

Query Match      1.7%; Score 17; DB 4; Length 289;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 943 aagagaagcagtgaa 959
Db 267 AAGAGAACGACGTGAA 251

```

```

RESULT 7
US-08-184-009-110
; Sequence 110, Application US/08184009
; Patent No. 5833975
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo

```

APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/184,009
FILING DATE: 19-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 1084 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-184-009-110

Query Match 1.7%; Score 17; DB 2; Length 1084;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 463 ggtcctgggaccctcg 479
|||||
DB 292 ggtcctgggaccctcg 308

RESULT 8
US-08-458-356-110
Sequence 110, Application US/08458356
Patent No. 5942235
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,356
FILING DATE: 02-JUN-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/184,009
FILING DATE: 19-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 1084 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-458-356-110

Query Match 1.7%; Score 17; DB 2; Length 1084;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 463 ggtcctgggaccctcg 479
|||||
DB 292 ggtcctgggaccctcg 308

RESULT 9
US-08-184-009-109
Sequence 109, Application US/08184009
Patent No. 5833975
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/184,009
FILING DATE: 19-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 1094 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-184-009-109

Query Match 1.7%; Score 17; DB 2; Length 1094;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 ggtcctgggacccctgg 479
|||||
DB 314 GGTCTGTGGCACCCTGG 330

RESULT 10

US-08-458-356-109
Sequence 109, Application US/08458356
Patent No. 5942235

GENERAL INFORMATION:

APPLICANT: Paolletti, Enzo
APPLICANT: Taitaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,356
FILING DATE: 02-JUN-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/184,009
FILING DATE: 19-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-0712
TELEFAX: (212) 840-3333
TELEX: 425066CURTMS

INFORMATION FOR SEQ ID NO: 109:

SEQUENCE CHARACTERISTICS:
LENGTH: 1094 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-458-356-109

Query Match 1.7%; Score 17; DB 2; Length 1094;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 ggtcctgggacccctgg 479
|||||
DB 314 GGTCTGTGGCACCCTGG 330

RESULT 11

US-09-118-442-29/C

Sequence 29, Application US/09118442B
Patent No. 6197561
GENERAL INFORMATION:
APPLICANT: Martino-Catt, Susan J.
APPLICANT: Wang, Hongyu

APPLICANT: Beach, Larry R.

APPLICANT: Wang, Xun

APPLICANT: Bowen, Benjamin A.

TITLE OF INVENTION: Genes Controlling Phytate Metabolism in

FILE REFERENCE: 0706

CURRENT APPLICATION NUMBER: US/09/118,442B

EARLIER FILING DATE: 1998-07-17

EARLIER APPLICATION NUMBER: 60/055,446

EARLIER FILING DATE: 1997-08-11

EARLIER APPLICATION NUMBER: 60/055,526

EARLIER FILING DATE: 1997-08-08

EARLIER APPLICATION NUMBER: 60/053,944

EARLIER FILING DATE: 1997-07-28

NUMBER OF SEQ ID NOS: 31

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 29

LENGTH: 1330

TYPE: DNA

ORGANISM: Zea mays

US-09-118-442-29

Query Match 1.7%; Score 17; DB 4; Length 1330;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 catccgatgaagattc 350
|||||
DB 493 CATCCGATGAGATTTC 477

RESULT 12

US-08-993-118-8

Sequence 8, Application US/08993118

Patent No. 5997872

GENERAL INFORMATION:

APPLICANT: Lucas, Sophie;
APPLICANT: DE SMET, Charles;
APPLICANT: BOON-PALLEUR, Thierry
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,118
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/845,528
FILING DATE: April 25, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mary Anne Schofield
REGISTRATION NUMBER: 36,669
REFERENCE/DOCKET NUMBER: LUD 5455
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1691 base pairs

;; TYPE: nucleotides
;; STRANDEDNESS: single stranded
;; TOPOLOGY: linear
US-08-993-118-8

Query Match 1.7%; Score 17; DB 2; Length 1691;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 463 ggtcctgggacacctgg 479
|||||
Db 317 ggtcctgggacacctgg 333

RESULT 13
US-08-845-528C-8
; Sequence 8, Application US/08845528C
; Patent No. 6027924
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie;
; APPLICANT: DE SMET, Charles;
; APPLICANT: BOON-FALLEUR, Thierry;
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGC-CI AND USES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,528C
; FILING DATE: April 25, 1997
; CLASSIFICATION: 4335
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1691 base pairs
; TYPE: nucleotides
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
US-08-845-528C-8

Query Match 1.7%; Score 17; DB 3; Length 1691;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 463 ggtcctgggacacctgg 479
|||||
Db 317 ggtcctgggacacctgg 333

RESULT 14
US-07-807-043B-7
; Sequence 7, Application US/07807043B
; Patent No. 5342774
; GENERAL INFORMATION:

;; APPLICANT: Boon, Thierry, Van den Eynde, Beno t
;; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
;; TITLE OF INVENTION: Rejection Antigen Antigen Precursors, Tumor
;; NUMBER OF SEQUENCES: 16
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Felfe & Lynch
;; STREET: 805 Third Avenue
;; CITY: New York City
;; STATE: New York
;; ZIP: 10022

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,043B
; FILING DATE: 19911212
; CLASSIFICATION: 424

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5342774man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 253.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2419 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: singular
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-07-807-043B-7

Query Match 1.7%; Score 17; DB 1; Length 2419;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 463 ggtcctgggacacctgg 479
|||||
Db 739 ggtcctgggacacctgg 755

RESULT 15
US-08-299-849B-7
; Sequence 7, Application US/08299849B
; Patent No. 5612201
; GENERAL INFORMATION:

;; APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
;; APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;
;; APPLICANT: Chomez, Patrick
;; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
;; TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
;; NUMBER OF SEQUENCES: 48
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Felfe & Lynch
;; STREET: 805 Third Avenue
;; CITY: New York City
;; STATE: New York
;; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

```
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,849B
FILING DATE: 1-SEPTEMBER-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/037,230
FILING DATE: 26-MARCH-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
APPLICATION NUMBER: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5612201man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5355
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ. ID NO. 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-299-849B-7

Query Match      1.7%; Score 17; DB 1; Length 2419;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 gtcctcggcaccctgg 479
|||||
Db 739 ggcctcggcaccctgg 755

RESULT 16
US-08-142-368A-7
Sequence 7, Application US/08142368A
Patent No. 5925729
GENERAL INFORMATION:
APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
TITLE OF INVENTION: Rejection Antigens and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
```

```
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/142,368A
FILING DATE: 02-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
APPLICATION NUMBER: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5925729man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5253.4-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ. ID NO. 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-142-368A-7

Query Match      1.7%; Score 17; DB 2; Length 2419;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 gtcctcggcaccctgg 479
|||||
Db 739 ggcctcggcaccctgg 755

RESULT 17
US-08-967-727-7
Sequence 7, Application US/08967727
Patent No. 6025474
GENERAL INFORMATION:
APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
APPLICANT: Van der Bruggen, Pierre; Boon-Falleur, Thierry
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,727
FILING DATE: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/037,230
FILING DATE: 26-MARCH-1993
```

APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,365
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6025474man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5353
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-967-727-7

Query Match 1.7%; Score 17; DB 3; Length 2419;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 ggtcctgggaccctgg 479
|||||
Db 739 ggtcctgggaccctgg 755

RESULT 18
US-08-037-230D-7
Sequence 7, Application US/08037230D
Patent No. 6235525
GENERAL INFORMATION:
APPLICANT: Gaugler, B alrice; Van den Eynde, Beno t;
APPLICANT: van der Bruggen, Pierre; Boon-Palleur, Thierry
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
NUMBER OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/037,230D
FILING DATE: 26-MARCH-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364

FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6235525man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5353
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-037-230D-7

Query Match 1.7%; Score 17; DB 4; Length 2419;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 ggtcctgggaccctgg 479
|||||
Db 739 ggtcctgggaccctgg 755

RESULT 19
US-08-465-167A-23
Sequence 23, Application US/08465167A
Patent No. 5750395
GENERAL INFORMATION:
APPLICANT: Fikes, John D.
APPLICANT: Livingston, Brian D.
APPLICANT: Sette, Alessandro D.
APPLICANT: Sidney, John C.
TITLE OF INVENTION: DNA ENCODING MAGE-1 C-TERMINAL
TITLE OF INVENTION: IMMUNOGENIC PEPTIDES (as amended)
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 98111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,167A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/103,623
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14137-60-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:

LENGTH: 2420 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 626..1552
 US-08-465-167A-23

Query Match 1.7%; Score 17; DB 1; Length 2420;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 ggtcctgggacccctgg 479
 ||||||||||||||||
 Db 739 ggtcctgggacccctgg 755

RESULT 20
 US-08-441-430-1
 ; Sequence 1, Application US/08441430
 ; Patent No. 5681942
 ; GENERAL INFORMATION:
 ; APPLICANT: Buchwald, Manuel
 ; APPLICANT: Strathdee, Craig A.
 ; APPLICANT: Wevrick, Rachel
 ; TITLE OF INVENTION: Fanconi Anemia Type C Gene
 ; NUMBER OF SEQUENCES: 73
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Richard J. Polley, Esq.
 ; ADDRESSEE: Klarquist, Sparkman, Campbell, Leigh &
 ; ADDRESSEE: Whinston, LLP
 ; STREET: 121 S.W. Salmon, Suite 1600
 ; City: Portland
 ; STATE: Oregon
 ; COUNTRY: U.S.A.
 ; ZIP: 97204
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Disk, 3+-inch
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: MS DOS
 ; SOFTWARE: Wordperfect 5.1/ASCII Text File
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/441,430
 ; FILING DATE: May 15, 1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: U.S. 07/876,285
 ; FILING DATE: April 29, 1992
 ; APPLICATION NUMBER: U.S. 07/918,313
 ; FILING DATE: July 21, 1992
 ; APPLICATION NUMBER: U.S. 08/003,963
 ; FILING DATE: January 15, 1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Richard J. Polley, Esq.
 ; REGISTRATION NUMBER: 28,107
 ; REFERENCE/DOCKET NUMBER: 3812-42824
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (503) 226-7391
 ; TELEFAX: (503) 228-9446
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4488 base pairs
 ; TYPE: Nucleic Acid
 ; STRANDEDNESS: Double stranded
 ; TOPOLOGY: Linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:

ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 LIBRARY: Human cDNA
 POSITION IN GENOME: (of corresponding genomic gene)
 CHROMOSOME/SEGMENT: 9q
 MAP POSITION: 22.3
 UNITS:
 US-08-441-430-1

Query Match 1.7%; Score 17; DB 1; Length 4488;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 917 tgaagccaccctgggaag 933
 ||||||||||||||||
 Db 4189 tgaagccaccctgggaag 4205

RESULT 21
 US-07-807-043B-8
 ; Sequence 8, Application US/07807043B
 ; Patent No. 5342774
 ; GENERAL INFORMATION:
 ; APPLICANT: Boon, Thierly, Van den Eynde, Beno t
 ; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felle & Lynch
 ; STREET: 805 Third Avenue
 ; City: New York City
 ; STATE: New York
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
 ; COMPUTER: IBM
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/807,043B
 ; FILING DATE: 19911212
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/764,364
 ; FILING DATE: 23-SEPTEMBER-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/728,838
 ; FILING DATE: 9-JULY-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/705,702
 ; FILING DATE: 23-May-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 5342774man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 253.3
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 688-9200
 ; TELEFAX: (212) 838-3884
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5674 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: singular
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 ; FEATURE:
 ; NAME/KEY: MAGE-1 gene
 ; US-07-807-043B-8
 ; Query Match 1.7%; Score 17; DB 1; Length 5674;
 ; Best Local Similarity 100.0%; Pred. No. 34;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 463 ggtcctggcaccctgg 479
|||||
Db 3994 GGTCTGGGACCCCTGG 4010

RESULT 22

US-08-190-411A-1
; Sequence 1, Application US/08190411A
; Patent No. 5541104
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng; Stockert, Elisabeth;
; APPLICANT: Chen, Yachi; Garin-Chesa, Pilar; Rettig, Wolfgang J.;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry;
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES WHICH BIND TO
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR MAGE-1, RECOMBINANT MAGE-1,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,411A
; FILING DATE: 01-FEBRUARY-1994
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 037,230
; FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; APPLICATION NUMBER: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5541104man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5354
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: MAGE-1 gene
; US-08-190-411A-1

Query Match 1.7%; Score 17; DB 1; Length 5674;

Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 463 ggtcctggcaccctgg 479
|||||
Db 3994 GGTCTGGGACCCCTGG 4010

RESULT 23

US-08-299-849B-8
; Sequence 8, Application US/08299849B
; Patent No. 561201
; GENERAL INFORMATION:
; APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
; APPLICANT: Leith, Bernard; Szkora, Jean-Pierre; De Smet, Charles;
; APPLICANT: Chomez, Patrick
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
; TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,849B
; FILING DATE: 1-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; APPLICATION NUMBER: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 561201man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5355
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: MAGE-1 gene
; US-08-299-849B-8

Query Match 1.7%; Score 17; DB 1; Length 5674;
Best Local Similarity 100.0%; Pred. No. 34;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 ggtcctggcaccctgg 479
|||||

Db 3994 ggtcctggcaccctgg 4010

RESULT 24

US-08-560-024-1

; Sequence 1, Application US/08560024

; Patent No. 5843448

; GENERAL INFORMATION:

; APPLICANT: Chen, Yao-Tseung; Stockert, Elisabeth;

; APPLICANT: Chen, Yachi; Garin-Chesa, Pilar; Rettig, Wolfgang J.;

; APPLICANT: van der Bruggen, Pierre; Boon-Fallour, Thierry;

; APPLICANT: Old, Lloyd J.

; TITLE OF INVENTION: MONOCLONAL ANTIBODIES WHICH BIND TO

; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR MAGE-1, RECOMBINANT MAGE-1,

; TITLE OF INVENTION: AND MAGE-1 DERIVED IMMUNOGENIC PEPTIDES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felle & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/560,024

; FILING DATE:

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/190,411

; FILING DATE: 01-FEBRUARY-1994

; APPLICATION NUMBER: 037,230

; FILING DATE: 26-MARCH-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/04354

; FILING DATE: 22-MAY-1992

; APPLICATION NUMBER: 07/807,043

; FILING DATE: 12-DECEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/764,364

; FILING DATE: 23-SEPTEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/728,838

; FILING DATE: 9-JULY-1991

; APPLICATION NUMBER: 07/705,702

; FILING DATE: 23-MAY-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 5843448man D.

; REGISTRATION NUMBER: 30,946

; REFERENCE/DOCKET NUMBER: LUD 5354

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5674 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

; FEATURE:

; NAME/KEY: MAGE-1 gene

; US-08-560-024-1

Query Match 1.7%; Score 17; DB 2; Length 5674;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 ggtcctggcaccctgg 479
|||||

Db 3994 ggtcctggcaccctgg 4010

RESULT 25

US-08-142-368A-8

; Sequence 8, Application US/08142368A

; Patent No. 5925729

; GENERAL INFORMATION:

; APPLICANT: Boon-Fallour, Thierry; Van der Bruggen, Thierry;

; APPLICANT: van den Eynde, Beno t; Van pel, Aline; De plaen, Etienne;

; APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia

; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor

; TITLE OF INVENTION: Rejection Antigens and Uses Thereof

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felle & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/142,368A

; FILING DATE: 02-MAY-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/04354

; FILING DATE: 22-MAY-1992

; APPLICATION NUMBER: 07/807,043

; FILING DATE: 12-DECEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/764,364

; FILING DATE: 23-SEPTEMBER-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/728,838

; FILING DATE: 9-JULY-1991

; APPLICATION NUMBER: 07/705,702

; FILING DATE: 23-MAY-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 5925729man D.

; REGISTRATION NUMBER: 30,946

; REFERENCE/DOCKET NUMBER: LUD 5253.4-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5674 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

; FEATURE:

; NAME/KEY: MAGE-1 gene

; US-08-142-368A-8

Query Match 1.7%; Score 17; DB 2; Length 5674;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 ggtcctggcaccctgg 479
|||||
DB 3994 GGTCTGGGACCCCTGG 4010

RESULT 26
US-08-967-727-8
; Sequence 8, Application US/08967727
; Patent No. 6025474
; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,727
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,365
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6025474man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: MAGE-1 gene
; US-08-967-727-8

Query Match 1.7%; Score 17; DB 3; Length 5674;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 463 ggtcctggcaccctgg 479
|||||

DB 3994 GGTCTGGGACCCCTGG 4010

RESULT 27
US-08-037-230D-8
; Sequence 8, Application US/08037230D
; Patent No. 6235525
; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/037,230D
; FILING DATE: 26-MARCH-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6235525man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: MAGE-1 gene
; US-08-037-230D-8

Query Match 1.7%; Score 17; DB 4; Length 5674;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 463 ggtcctggcaccctgg 479
|||||
DB 3994 GGTCTGGGACCCCTGG 4010
RESULT 28
US-07-853-913-1

```
; Sequence 1, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; TITLE OF INVENTION: Neuroepithelial Tumors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853,913
; FILING DATE: 19920319
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/660,412
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,803
; FILING DATE: 25-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/201,762
; FILING DATE: 02-JUN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/180,548
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-6240
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 11236 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-07-853-913-1

Query Match      1.7%; Score 17; DB 1; Length 11236;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 689 cctcaggaaggagctg 705
      |||||
Db 3466 CCTCAGGAAGGGGCTG 3482
```

```
RESULT 29
US-09-335-409-1/c
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPORHILONES
```

```
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1
```

```
Query Match      1.7%; Score 17; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 268 cggccagcccccagcag 284
      |||||
Db 20332 CGGCCAGCCCCCAGCAG 20316
```

```
RESULT 30
US-07-753-110B-12
; Sequence 12, Application US/07753110B
; Patent No. 5436141
; GENERAL INFORMATION:
; APPLICANT: Miyata, Shohei
; APPLICANT: Ohshima, Atsushi
; APPLICANT: Inouye, Sumiko
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: METHOD FOR SYNTHESIZING STABLE
; TITLE OF INVENTION: SINGLE-STRANDED cDNA IN EUKARYOTES BY MEANS OF A BACTERIAL
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/753,110B
; FILING DATE: 30-AUG-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 377.5584P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19
; OTHER INFORMATION: /note= "The 2' position of the 5' position of
; OTHER INFORMATION: nucleotide is linked to the 5' position of
; OTHER INFORMATION: nucleotide number 1 of SEQ ID NO: 11 of this
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 69..76
```



```

ADDRESSER: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 207:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: other
LOCATION: 60..181
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 97
OTHER INFORMATION: region 1..122
OTHER INFORMATION: id AA057454
FEATURE:
NAME/KEY: other
LOCATION: 182..233
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 122..173
OTHER INFORMATION: id AA057454
FEATURE:
NAME/KEY: other
LOCATION: 71..233
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 1..163
OTHER INFORMATION: id C18312
FEATURE:
NAME/KEY: other
LOCATION: 182..233
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 144..195
OTHER INFORMATION: id W69247
FEATURE:
NAME/KEY: other
LOCATION: 98..144
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 62..108
OTHER INFORMATION: id W69247
FEATURE:

```

```

NAME/KEY: other
LOCATION: 34..78
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 97
OTHER INFORMATION: region 1..45
OTHER INFORMATION: id W69247
FEATURE:
NAME/KEY: other
LOCATION: 146..233
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 69..156
OTHER INFORMATION: id H75891
FEATURE:
NAME/KEY: other
LOCATION: 76..144
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 95
OTHER INFORMATION: region 1..69
OTHER INFORMATION: id H75891
FEATURE:
NAME/KEY: other
LOCATION: 80..233
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 1..154
OTHER INFORMATION: id HUM11265
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 104..160
IDENTIFICATION METHOD: Von Heljne matrix
OTHER INFORMATION: score 7.4
OTHER INFORMATION: seq PHILTRALQARRA/GP
US-08-905-223-207

Query Match 1.6%; Score 16; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 647 ccagagagtcctgcac 662
Db 171 CCAGAGGTCCTGCAC 156

RESULT 34
US-08-630-822A-97/c
; Sequence 97, Application US/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDIA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/630,822A
FILING DATE: 11-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 252 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-630-822A-97

Query Match 1.6%; Score 16; DB 2; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 aagaagtcagtaacca 304
|||||
Db 195 AAGAGTCACTACCA 180

RESULT 35
US-09-005-069-97/C
Sequence 97, Application US/09005069

Patent No. 5932470
GENERAL INFORMATION:

APPLICANT: FRANK, GLENN R.
APPLICANT: HUNTER, SHIRLEY WU
APPLICANT: WALLENFELDS, LYNDIA
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,069

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,822
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 252 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

US-09-005-069-97

Query Match 1.6%; Score 16; DB 2; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 aagaagtcagtaacca 304
|||||
Db 195 AAGAGTCACTACCA 180

RESULT 36
US-08-906-769-104/C
Sequence 104, Application US/08906769
Patent No. 6077687
GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Slegler, Gary
APPLICANT: Gaines, Patrick J.

APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,769

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:

LENGTH: 252 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 1..251
OTHER INFORMATION: /note="At pos. bp 4, change A to
OTHER INFORMATION: R. At pos. aa 2, substitute Xaa."
US-08-906-769-104

Query Match 1.6%; Score 16; DB 3; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 aagaagtcagtaacca 304

Db 199 AAGAAGTCAGTACCCA 184

```
RESULT 37
US-08-906-616-104/C
; Sequence 104, Application US/08906616
; Patent No. 6121035
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,616
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..251
; OTHER INFORMATION: /note= "At pos. bp 4, change A to
; OTHER INFORMATION: R. At pos. aa 2, substitute Xaa."
US-08-906-616-104

Query Match 1.6%; Score 16; DB 3; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 aagaagtcagtagcca 304
Db 199 AAGAAGTCAGTACCCA 184

RESULT 38
US-08-817-795-104/C
; Sequence 104, Application US/08817795
; Patent No. 6139840
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Hunter, Shirley Wu
```

```
; APPLICANT: Frank, Glenn R.
; APPLICANT: Heath, Andrew W.
; APPLICANT: Yamaka, Miles Yamataka
; APPLICANT: Aristen, Ann
; APPLICANT: Dale, Beverly
; APPLICANT: Stiegler, Gary
; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
; TITLE OF INVENTION: PROPRIATE VACCINES TO PROTECT ANIMALS FROM FLEA
; TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,795
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14442
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gary J. Connell
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..251
; NAME/KEY: Xaa = any amino acid
; LOCATION: 2
US-08-817-795-104

Query Match 1.6%; Score 16; DB 3; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 aagaagtcagtagcca 304
Db 199 AAGAAGTCAGTACCCA 184

RESULT 39
US-08-639-075A-104/C
; Sequence 104, Application US/08639075A
; Patent No. 6150125
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
```

APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,075A
FILING DATE: 24-APR-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 252 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..251
OTHER INFORMATION: /note= "At pos. bp 4, change A to
OTHER INFORMATION: R. At pos. aa 2, substitute Xaa."
US-08-639-075A-104

Query Match 1.6%; Score 16; DB 3; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 289 aagaagtcagtcacca 304
|||||
DB 199 AAGAAGTCAGTACCCA 184

RESULT 40
US-09-012-431-104/C
Sequence 104, Application US/09012431
Patent No. 6180383
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
Rushlow, Keith E.
Wu Hunter, Shirley
Frank, Glenn R.
Stiegler, Gary
Gaines, Patrick J.
Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,431
FILING DATE: 23-Jan-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 252 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..251
OTHER INFORMATION: /note= "At pos. bp 4, change A to
R. At pos. aa 2, substitute Xaa."
SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-09-012-431-104

Query Match 1.6%; Score 16; DB 4; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 289 aagaagtcagtcacca 304
|||||
DB 199 AAGAAGTCAGTACCCA 184

RESULT 41
US-09-012-692-104/C
Sequence 104, Application US/09012692
Patent No. 6214579
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
Rushlow, Keith E.
Wu Hunter, Shirley
Frank, Glenn R.
Stiegler, Gary
Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,692


```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..251
; OTHER INFORMATION: /note= "At pos. bp 4, change A to
; OTHER INFORMATION: R. At pos. aa 2, substitute Xaa."
US-09-012-692-104

Query Match      1.6%; Score 16; DB 4; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 aagaagtcagttaccga 304
Db 199 AAGAAGTCAGTACC GA 184

RESULT 42
US-08-906-613-104/C
; Sequence 104, Application US/08906613
; Patent No. 6232096
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
```

```

; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..251
; OTHER INFORMATION: /note= "At pos. bp 4, change A to
; OTHER INFORMATION: R. At pos. aa 2, substitute Xaa."
US-08-906-613-104

Query Match      1.6%; Score 16; DB 4; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 aagaagtcagttaccga 304
Db 199 AAGAAGTCAGTACC GA 184

RESULT 43
PCT-US95-14442A-104/C
; Sequence 104, Application PC/TUS9514442A
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Frank, Glenn R.
; APPLICANT: Heath, Andrew W.
; APPLICANT: Yamaka, Miles Yamanaka
; APPLICANT: Aristen, Ann
; APPLICANT: Dale, Beverly
; APPLICANT: Stiegler, Gary
; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
; TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14442A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gary J. Connell
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 base pairs
; TYPE: nucleic acid
```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..251
; NAME/KEY: Xaa = any amino acid
; LOCATION: 2
; PCT-US95-14442A-104

```

```

Query Match 1.6%; Score 16; DB 5; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 289 aagaatcagtcagccca 304
Db 199 AAGAGTCTAGTACCCA 184

```

```

RESULT 44
US-09-060-756-254/C
; Sequence 254, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 254
; LENGTH: 291
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-254

```

```

Query Match 1.6%; Score 16; DB 4; Length 291;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 61 caccagttccgctgt 76
Db 83 CACCAgTTCGCTGT 68

```

```

RESULT 45
US-08-334-254-7
; Sequence 7, Application US/08334254
; Patent No. 5723290
; GENERAL INFORMATION:
; APPLICANT: James Eberwine, Marc Dichter, Kevin Miyashiro
; TITLE OF INVENTION: USE OF NEURITE LOCALIZED MRNAS FOR
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jane Massey Licata, Esq.
; STREET: 210 Lake Drive East, Suite 201
; CITY: Cherry Hill
; STATE: NJ
; COUNTRY: USA

```

```

; ZIP: 08002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/334,254
; FILING DATE: Herewith
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane Massey Licata
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: PENN-0028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 779-2400
; TELEFAX: (609) 779-8488
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; ANTI-SENSE: NO
US-08-334-254-7

```

```

Query Match 1.6%; Score 16; DB 1; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 693 caggaaaggagcgcgaga 708
Db 184 CAGGAAGGGGCTGAGA 199

```

Search completed: October 28, 2001, 22:34:10
Job time: 2648 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 21:23:22 ; Search time 1154.11 Seconds
(without alignments)
8198.791 Million cell updates/sec

Title: SEOLINS_COPY_1140_2140

Perfect score: 1001
Sequence: 1 tcgcagccacacagattcaaa.....aagccatcagcgtgggatg 1001

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 segs, 4726426750 residues

Word size : 8

Total number of hits satisfying chosen parameters: 19872988

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```
EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estrov1:*
71: em_estrov2:*
72: em_estp11:*
73: em_estp12:*
74: em_estp13:*
75: em_estp14:*
76: em_estp15:*
77: em_estp16:*
78: em_estp17:*
79: em_estp18:*
80: em_estp19:*
81: em_estp110:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*
```

117: gb_est148:*
 118: gb_est149:*
 119: gb_est150:*
 120: gb_est151:*
 121: gb_est152:*
 122: gb_est153:*
 123: gb_est154:*
 124: gb_est155:*
 125: gb_est156:*
 126: gb_est157:*
 127: gb_est158:*
 128: gb_est159:*
 129: gb_est160:*
 130: gb_est161:*
 131: gb_est162:*
 132: gb_est163:*
 133: gb_est164:*
 134: gb_est165:*
 135: gb_est166:*
 136: gb_est167:*
 137: gb_est168:*
 138: gb_est169:*
 139: gb_est170:*
 140: gb_est171:*
 141: gb_est172:*
 142: gb_est173:*
 143: gb_est174:*
 144: gb_est175:*
 145: gb_est176:*
 146: gb_est177:*
 147: gb_est178:*
 148: gb_est179:*
 149: gb_est180:*
 150: gb_est181:*
 151: gb_est182:*
 152: gb_est183:*
 153: gb_est184:*
 154: gb_est185:*
 155: gb_est186:*
 156: gb_est187:*
 157: gb_est188:*
 158: gb_est189:*
 159: gb_est190:*
 160: gb_est191:*
 161: gb_est192:*
 162: gb_est193:*
 163: gb_est194:*
 164: gb_est195:*
 165: gb_est196:*
 166: gb_est197:*
 167: gb_est198:*
 168: gb_est199:*
 169: gb_est200:*
 170: gb_est201:*
 171: gb_est202:*
 172: gb_est203:*
 173: gb_est204:*
 174: gb_est205:*
 175: gb_est206:*
 176: gb_est207:*
 177: gb_est208:*
 178: gb_est209:*
 179: gb_est210:*
 180: gb_est211:*
 181: gb_est212:*
 182: gb_est213:*
 183: gb_est214:*
 184: gb_est215:*
 185: gb_est216:*
 186: gb_est217:*
 187: gb_est218:*
 188: gb_est219:*
 189: gb_est220:*

190: gb_est110:*
 191: gb_est111:*
 192: gb_est112:*
 193: gb_est113:*
 194: gb_est114:*
 195: gb_est115:*
 196: gb_est116:*
 197: gb_est117:*
 198: gb_est118:*
 199: gb_est119:*
 200: gb_est120:*
 201: gb_est121:*
 202: gb_est122:*
 203: gb_est123:*
 204: gb_est124:*
 205: gb_est125:*
 206: gb_est126:*
 207: gb_est127:*
 208: gb_est128:*
 209: gb_est129:*
 210: gb_est130:*
 211: gb_est131:*
 212: gb_est132:*
 213: gb_est133:*
 214: gb_est134:*
 215: gb_est135:*
 216: gb_est136:*
 217: gb_est137:*
 218: gb_est138:*
 219: gb_est139:*
 220: gb_est140:*
 221: gb_est141:*
 222: gb_est142:*
 223: gb_est143:*
 224: gb_est144:*
 225: gb_est145:*
 226: gb_est146:*
 227: gb_est147:*
 228: gb_est148:*
 229: gb_est149:*
 230: gb_est150:*
 231: gb_est151:*
 232: gb_est152:*
 233: gb_est153:*
 234: gb_est154:*
 235: gb_est155:*
 236: gb_est156:*
 237: gb_est157:*
 238: gb_est158:*
 239: gb_est159:*
 240: gb_est160:*
 241: gb_est161:*
 242: gb_est162:*
 243: gb_est163:*
 244: gb_est164:*
 245: gb_est165:*
 246: gb_est166:*
 247: gb_est167:*
 248: gb_est168:*
 249: gb_est169:*
 250: gb_est170:*
 251: gb_est171:*
 252: gb_est172:*
 253: gb_est173:*
 254: gb_est174:*
 255: gb_est175:*
 256: gb_est176:*
 257: gb_est177:*
 258: gb_est178:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	509	50.8	676	165	BE260495	BE260495 601150702
2	502	50.1	664	166	BE383336	BE383336 601298249
3	470	47.0	692	166	BE382353	BE382353 601298656
4	452	45.2	823	158	BE619259	BE619259 601473130
5	410	41.0	614	152	BE335963	BE335963 602404708
6	399	39.9	960	169	BE797306	BE797306 602256926
7	381	38.1	915	147	BE311926	BE311926 601897767
8	378	37.8	761	140	BE795820	BE795820 601590856
9	371	37.1	884	154	BE519751	BE519751 602578902
10	355	35.5	745	141	BE900936	BE900936 601674206
11	354	35.4	1012	153	BE386348	BE386348 602455550
12	337	33.7	920	153	BE386090	BE386090 602455264
13	330	33.0	992	139	BE747163	BE747163 601577254
14	324	32.4	878	147	BE338283	BE338283 602035507
15	317	31.7	574	165	BE250309	BE250309 600943455
16	313	31.3	762	107	AU124671	AU124671 AU124671
17	312	31.2	1497	147	BE315158	BE315158 601899518
18	307	30.7	735	141	BE902696	BE902696 601677393
19	289	28.9	431	5	AA310236	AA310236 EST181085
20	289	28.9	612	166	BE304720	BE304720 601106236
21	282	28.2	307	188	T34216	T34216 EST64346 Hu
22	280	28.0	947	150	BE525432	BE525432 602069517
23	279	27.9	1124	147	BE342802	BE342802 602015077
24	271	27.1	677	139	BE742908	BE742908 601574609
25	262	26.2	482	187	R55841	R55841 YG89401.r1
26	253	25.3	944	172	BP696043	BP696043 602269966
27	247	24.7	517	104	AI991599	AI991599 WS18G04.x
28	246	24.6	446	10	AA634909	AA634909 ab27f02.r
29	243	24.3	346	122	AM889463	AM889463 RC6-NT002
30	240	24.0	501	5	AA311855	AA311855 EST182568
31	230	23.0	698	140	BE795434	BE795434 601592991
32	223	22.3	938	165	BE260626	BE260626 601146116
33	209	20.9	278	120	AM805551	AM805551 IL0-ST000
34	209	20.9	282	188	T34024	T34024 EST61387 Hu
35	206	20.6	963	154	BE480926	BE480926 602530075
36	199	19.9	472	187	R51138	R51138 YG71C08.r1
37	197	19.7	891	175	BE258460	BE258460 602379938
38	195	19.5	906	152	BE336190	BE336190 602404980
39	195	19.5	961	153	BE387104	BE387104 602453813
40	185	18.5	394	6	AA346268	AA346268 EST52407
41	185	18.5	936	174	BE116283	BE116283 602318546
42	182	18.2	547	168	BE686235	BE686235 602143687
43	179	17.9	688	140	BE794311	BE794311 601591442
44	175	17.5	852	146	BE240253	BE240253 601905776
45	174	17.4	441	157	H03317	H03317 YJ47e10.r1

ALIGNMENTS

RESULT	1					
BE260495		676 bp	mrna	EST	26-OCT-2000	
LOCUS	60115070221	NIH_MGC_19	Homo sapiens	cdna clone	IMAGE:3503184.5	
DEFINITION	mrna sequence.					
ACCESSION	BE260495					
VERSION	BE260495.1	GI:9131807				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
AUTHORS	1 (bases 1 to 676)					
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .					
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)					
COMMENT	Unpublished (1999)					
	Contact: Robert Strausberg, Ph.D.					

Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: L10M176 row: d column: 01
High quality sequence stop: 672.
Location/Qualifiers

FEATURES

source

1. 676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3503184"
/clone_1lb="NIH-MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCCGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

BASE COUNT 154 a 207 c 176 g 139 t
ORIGIN

Query Match

50.8%; Score 509; DB 165; Length 676;

Best local Similarity 99.8%; Pred. No. 1,1e-259;

Matches 62; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	1	tcgcagcacaagaatccaacccagctcaacccatccacccggacacatctccctcgt	60
DB	48	TCGCAGCCACAAAGATCAAAACCCAGCTCAACCTCATCCACCCGACATCTCCCTGCT	107
QY	61	caccagcttcgcgtgaagaaggagcccccacccctcaagtgtgcctatgttcagggtga	120
DB	108	CACCAAGTTCCGCTGAAGAGAGGAGGCCCCCACCCTCAGTGTGCCATGTTCAGGGTGA	167
QY	121	atgcctctcaagtaagcagctccgcccagaaggaggtgagagggatgcctattac	180
DB	168	ATGCCTCTCAAGTACCAAGCTCCCTCCAGAGGAGGATGCGAGGATGCGATTATTAC	227
QY	181	ttgcaatccctgaggaatcatagttgagcgctgcaagcttcccaactccagcagagcgt	240
DB	228	TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCTGCTCCCAACTCCAGCAGAGCGT	287
QY	241	gcagagtaacagagagagtgagcagagagcagcccccagcagaagaagaatcagta	300
DB	288	GCAGAGTACAGGAGGAGTGCGCAGAGCGCCGCCACCAAGAAAGAAAGTCAAGTA	347
QY	301	cccagaatacatctctcttgaacagaggtctgcacatccagatgaagattcgaatgtcag	360
DB	348	CCCAAGAAATCATCTTCTTGGAAACAGGCTTCCCATCCCAATGAAGATTGCAAAATGTCAG	407
QY	361	ttccacactgttcaacalaagcccccagacagctctctgtactatgagctgtgtgagggcac	420
DB	408	TGCCACACTGTCAACAATAAGCCCGGACAGCTCTGTGTACTGACAGTGTGTGAGGGGAC	467
QY	421	attggagcagctgtgcgctcatcaggaagacaggtgtgacagggctcctggcagccctggc	480
DB	468	ATTGGGACAGCTGTGCCGTCTATTACGAGACAGTGTGAGGAGGCTCTGGGACCCCTGGC	527
QY	481	tgctgtgtgtgtgtccacccctggcagcagagatcacacacaggggttcccaagtattcgc	540
DB	528	TGCTGTGTGTGTGTCCACCT-GCACCGAGATCACACACGGGCTTGCCAGATTTCTTGC	586
QY	541	ttcagagaagaacgccttggacatcttgggaaagcgccttacccttggctgtgtgtg	600
DB	587	TGCAGAGAGAAAGCGCCTTGCGATCTTTGGGAAAGCGGCTTCACCTTTGCTGTGGTGG	646

```
QY 601 cccccaacagctcaaacgctgagctccagc 630
DB 647 cccccaacagctcaaacgctgagctccagc 676

RESULT 2
BE383336 664 bp mRNA EST 21-JUL-2000
LOCUS 601298249F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5',
DEFINITION mRNA sequence.
ACCESSION BE383336
VERSION BE383336.1 GI:9328701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 664)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LNCM313 row: a column: 13
High quality sequence stop: 662.
Location/Qualifiers
1.664
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3628308"
/clone_1ib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 154 a 199 c 174 g 137 t
ORIGIN

Query Match 50.1%; Score 502; DB 166; Length 664;
Best Local Similarity 100.0%; Pred. No. 6e-256;
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcagcacaagatcaaacacagctcaacctcaccacccggagacatctccctctgt 60
DB 48 TCGACGCCCAAGATTCACAACCCAGCTCAACCTCATCCACCCGGACATCTCCCTCGT 107
QY 61 caccagtttcgctgtgaagaagagggggcccaacctcaatgtgcccataagtgatga 120
DB 108 CACCAATTTCCGCTGTAAAGAGAGGGCCCAACCCCACTGTGCCATGTTCAGGGTGA 167
QY 121 atgcgcctccatgacagctccgctcccgaggagagtgccagagagatgcatattac 180
DB 168 ATGCCCTCCCAAGTACACACTCCGTCACAGAGAGAGTGGCAAGGATGCCATTATTA 227
QY 181 ttgcgaatcccttggaattcatagttgagggcgctgcagcttcccaattccagagaagc 240
DB 228 TTGCAATCCTGAGGAATTCATAGTTGAGCGCGTCCCAATTCACAGAGAGCGT 287
QY 241 gcagagagtaacagagagatgagcagagagcgccacgagcagagagaagaagtaagta 300
|||||
```

```
DB 288 GCAGAGATACAGAGAGAGTGCAGGAGACGCCACGCCACAGAGAAAAAGATCAGTA 347
QY 301 cccaagaatcatcttcccttggaacagaggtctgcacatcccatgaagattcgaaatgcg 360
DB 348 CCCAGAAATCATCTCTCTTGGAAACAGAGGTGTGCATCCGATGAATTCGAAATGTCAG 407
QY 361 tgcacaactgtcaacataagccccgcacacgctctcgtcacttgcagctgtgtgagggcac 420
DB 408 TGGCACAATCTGTCAATATTAAGCCCGACACAGTCTCTGCTACTGTGACTGTGAGGGCAC 467
QY 421 atttgagcagctgtgctgcgtcatcattcagagagcaggtgcagaggtctctggaccctgc 480
DB 468 ATTTGGGAGCATGTGTGCCCTCATTAACGAGACGAGTGCAGAGGGTCTCGGCACCTTGC 527
QY 481 tgcgtgtgtgtgtccacctg 502
DB 528 TGCTGTGTGTGTGTCCACCTG 549

RESULT 3
BE382353 692 bp mRNA EST 21-JUL-2000
LOCUS 601298656F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3629028 5',
DEFINITION mRNA sequence.
ACCESSION BE382353
VERSION BE382353.1 GI:9327718
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 692)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LNCM314 row: o column: 13
High quality sequence stop: 600.
Location/Qualifiers
1.692
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3629028"
/clone_1ib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 152 a 193 c 197 g 150 t
ORIGIN

Query Match 47.0%; Score 470; DB 166; Length 692;
Best Local Similarity 99.8%; Pred. No. 6.7e-239;
Matches 590; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 336 tccgatgaagattcgaaatgtcagtcacacactgttaacataagccccgagacagtttc 395
DB 1 TCCGATGAAGATTCGAAATGTCACTGCACACTTGTCAACATAAGCCCGACACAGCTCTC 60
|||||
```


BASE COUNT	185 a	219 c	259 g	160 t					
ORIGIN	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 Kb. Library constructed by Life Technologies.								
Query Match	45.2%; Score 452; DB 138; Length 823;								
Best Local Similarity	100.0%; Pred. No. 2,66-229;								
Matches 452;	Conservative	0;	Mismatches	0;	Indels 0; Gaps 0				
QY	503	gcacgcagatcaaccacacgagcgttcccaagtatctgtcgtcagagagaagcgccttgc	562						
Db	88	GCACGCGAATCACACACACGGGCTTGGCAAGTACTTGTGTCAGAGAAACGGGCTTGGC	147						
OY	563	atcttttggaaagccggttcaaccttgcgtggtgtgtccccaacacagctcaaaagcgt	622						
Db	148	ATCTTTTGGGAAAGCCGCTTACCCCTTTTGTCTGTGTGTGGTGGCCCAACACAGCTCAAGGCTG	207						
OY	623	gtctcagcagtacccacaacacagtcgccaaggaagtcctctgcacacacatcatgatctctgc	682						
Db	208	GCTCCAGCAGTACCAACACAGTGCACGAGGTCCTGCACACATCAGTATGATTCCTGC	267						
OY	683	caaatgccttcaggaagggcctgagatctccagctcctcagtcagtcgaaagatgatcagtc	742						
Db	268	CAAAATGCTTTCAGGAAGGGGCTGAGATCTCCATCTCCTGCAGTGAAGAAATGATCAGTTC	327						
OY	743	gcttttcggaacatgtgattttggaagaatttcagacctctcgtgtgctggagccttgaaaca	802						
Db	328	GCTGTTCGGAACATGTGATTTGGAAAGTTTCAGACCTTCTGTGTGGCGCACTGCACGA	387						
OY	803	tgcgcttgcgtctgcgtctgtgtcacacctctgcgttgaaagtcgtctatcccgaggacac	862						
Db	388	TGGCTTTTGGCTGTGCGTGTGTGCACACCTGTGGCTTGGAAGTGCTATTCGGGGACAC	447						
OY	863	catgcacctgcagagcctctgtctccgagatggggaaagaatgcccacctctgatatcgaagc	922						
Db	448	CATGCCCTGGGAGGCTCTGCTGCCGATGGGGAAAGATGCCACCTCTCATGATCATGAAGC	507						
OY	923	caccttggaagatgtgtttggaagaagaagcag	954						
Db	508	CACCTTGAAGATGTGTTTGGAAAGAGAGACGAG	539						
RESULT	5								
LOCUS	BG335963	614 bp	mRNA	EST	27-FEB-2001				
DEFINITION	6020404708F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4542553 5', mRNA sequence.								
ACCESSION	BG335963								
VERSION	1								
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.								
AUTHORS	1 (bases 1 to 614)								
TITLE	NIH-MGC http://mgc.nci.nih.gov/								
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)								
COMMENT	Unpublished (1999)								
	Contact: Robert Strassberg, Ph.D.								
	Email: cgaabs-r@mail.nih.gov								
	Tissue Procurement: ATCC								
	CDNA Library Preparation: Ling Hong/Rudin Laboratory								
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)								
	DNA Sequencing by: Incyte Genomics, Inc.								
	Clone distribution: MGC clone distribution information can be								
	found through the I.M.A.G.E. Consortium/LLNL at:								
	http://image.llnl.gov								
	plate: L1CM1222 row: 0 column: 02								
	High quality sequence stop: 614.								
FEATURES									
source	1..614								
	Location/Qualifiers								

TITLE	JOURNAL	COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)	
Contact: Robert Strausberg, Ph.D.		
Email: cgaaps-remail.nih.gov		
Tissue Procurement: Louis Staudt, M.D., Ph.D.		
cDNA Library Preparation: Life Technologies, Inc.		
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
DNA Sequencing by: Incyte Genomics, Inc.		
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov		
Plate: ILAM9952	row: c	column: 07
High quality sequence stop: 705.		
location/qualifiers		

```
source      1. .960
            /organism="Homo sapiens"
```

```

/clone_id="NH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"

```

Site_2: SalI; cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

Query Match	39.9%;	Score 399;	DB 169;	Length 960;
Best Local Similarity	99.4%;	Pred. No. 4.4e-201;		
Matches 669; Conservative	0;	Mismatches 3;	Indels 1;	Gaps 1

Db 1 CAACCTCCAGCAGAGCGTGCAGAGACTACAGAGAGAGTGCAGAGAGGCCAGGCCACAGC 60

QY 343 gaagattcgaaatgctcagtcgcaacattgtcaacataagccccgacacgctctctgctact 402
|||||
Db 121 GAAGATTCGAATGTCAGTCGCCACCTTGCAACATTAAGCCCCGACACGTCCTGCTACT 180

Db 181 GGACTGTGGTGAGGGCACGCTTTGGGACAGCTGTGCCCTATTACGGAGACCAAGTGGACAG 240

QY 523 gcttgcgaagtatctctgtcgcgagagagaacgcgccttgcactcttgygaaagccgcttc 582
|||||
Db 300 GCTTGCCTCAAGTATCTTGTCTGCACAGAGAGAACGCGCCTTGGCTCTCTTTGGGGAAGCCGCTTC 359
|||||

Db 360 ACCCTTTGCTGGTGTGCCCCAACCGCTCAAGCCTGGCTCCAGCAGTACCAAC 419
|||||
643 aatgccagaatcctgccaccatcattatattcctgccaatgacctcaagaaagga 702

Qy 703 ctgaagatctccagtcctcgtgcagtgtgaaagaattgatcatcagttcgctggttgcgaacatgttatt 762
|||||
Dh 480 cttcaactatgccccaaccggcggcacaggacaacagcttgctgacgggggcccgaataatgaccttc 539
|||||

Db 540 TGGAGAGTTTCAGACCTGTCTGGTGGCGCACTGCAGCATGCCGATTGGCTGTGCGCTGG 599

QY	823	ttgcacctctggcctgggaagatggtcattccggagggaacacatgccttcgagagctctgg	882
Db	600	TCGACACCTCTGGCTGGAAAGTGCTATTCCGGGACACCATGCCCTCGAGGCTCTGG	659
QY	883	tcgcgagatgaggaa 895	
Db	660	TCCGATGGGAA 672	
RESULT	7		
LOCUS	BF311926		
DEFINITION	BF311926 915 bp mRNA	EST	21-NOV-2000
ACCESSION	60189776761 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4126864 5',		
VERSION	mRNA sequence.		
KEYWORDS	BF311926		
SOURCE	BF311926.1 GI:11259697		
ORGANISM	EST.		
REFERENCE	human.		
AUTHORS	Homo sapiens		
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
COMMENT	NIH-MGC http://mhc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgaps@email.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: Image.lnl.gov Plate: LNCM017 row: I column: 17 High quality sequence stop: 672. Location/Qualifiers 1. 915 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4126864" /clone_lib="NIH_MGC_19" /tissue_type="neuroblastoma" /lab_host="DH10B (phage-resistant)" /note="Organ: brain; Vector: pORF7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5'- adapter: GGCACAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."		
FEATURES			
source			
BASE COUNT	206 a 262 c 286 g 159 t		
ORIGIN			
Query Match	38.1%; Score 381; DB 147; Length 915;		
Best Local Similarity	99.8%; Pred. No. 1.7e-191;		
Matches 501; Conservative 0; Mismatches 0; Indels 1; Gaps 1			
QY	1	tcgcagcacacaagatctcaaacccagctccaccatcatcacccggacatctccctgt 60	
Db	68	TCGCAGCGACACAAGATCAAAACCCAGCTCAACCTCATATCCACCGACATCTTCCCTGCT 127	
QY	61	caccagtttcgcgtgtagaagagagggccccacccttcagttgtgccatgttcagggtga 120	
Db	128	CACCAAGTTTCGGCTGTAAAGAGAGGGGCCACCCACCTCAAGTGTCCCATGGTTCAGGGTGA 187	
QY	121	atgcctctcaagatcacagctcgcgtcccaaggagagtgccagagggatgcattatcac 180	
Db	188	ATGCGCTCTCAAGTACCAAGTCCGCTCCGACAGAGGAGTGGCANAAGGATGCCATTATTAC 247	
QY	181	ttgcaatctgtagaagatcatagtttagggcgtcgtcagccttcccaacttccagcagagcgt 240	

Db	Query Match	Best Local Similarity	Score	DB	Length
248	TTGAAATCCGAGGAATTCATAGTGGAGGGCGTCGACGCTTCCACACTCCAGCAGACGG	99.8%	37.8%	378	761
Qy	gcaagaglacagagagagctgcagacagccagccagccagcagagaaaagtcagta				
Db	GCAGAGATACAGAGGAGTGCGCAGCAGCG-CCAGGCCCGACAGAGAAAAGAGTCAGTA				
Qy	cccaagaatattctctcttggaacagggtgtgcattcccgatgaagaattccgaattcag				
Db	CCCGAAGATATCTCTCTTGGAACAGGGGTGTGCATCCGATCCGATGAAGATTCAG				
Qy	tgccacctgtcaacaataagccccagacagctctctcactgtgactgtgtgtagggc				
Db	TGCCACACTTGTCAACATTAAGCCCCGACAGCTCTCTCTACTGTGACTGTGTAGGGCAC				
Qy	attgggcagctgtgccgtcattacagagacacaggtgagacaggtctctggcaccctggc				
Db	ATTGGGAGAGCTGTGCCGTATTATCGAGACACAGGTGAGACAGGGTCTGGGCACCTGGC				
Qy	tgctgtgttgtgtgtccacctg				
Db	TGCTGTGTGTGTGTGTCCACCTG				
RESULT	8				
LOCUS	BE795820	761 bp	mrna	EST	20-SEP-2000
DEFINITION	601590850F1 NIH_MGC_7 Homo sapiens CDNA clone IMAGE:3945085 5',				
ACCESSION	BE795820				
VERSION	BE795820.1	GI:10217018			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryotes: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.				
AUTHORS	1 (bases 1 to 761)				
TITLE	NIH-MGC http://mgc.ncl.nih.gov/				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: DCTD/DTP CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov Plate: LINC802 row: p column: 14 High quality sequence stop: 761. location/Qualifiers 1. 761 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="3945085" /clone_id="NIH_MGC-7" /tissue_type="small cell carcinoma" /cell_line="MGC3" /lab_host="DH10B (phage-resistant)" /note="Organ: lung; Vector: pOT8; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(g). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."				
BASE COUNT	172 a	206 c	219 g	164 t	
ORIGIN					

Matches	498;	Conservative	0;	Mismatches	0;	Indels	1;	Gaps	1;
QY	503	gaacgcagatcaccacacgagcttgcgaaglatcttgcgaagaagacgcgcttgcg							
Db	163	GCACGAGATTCACACACAGGGCTTGGCAGTATCTGTGCACAGAGAAAGCCCTTGGC							
QY	563	atccttgggaagaacgcgcttgcgaagcttgcgtgtgtgttgcgcacacagctcaaaccttg							
Db	223	ACTCTTGGGAAAGCCCTTACACCTTGTGCTGGTGGCCCAACACAGCTCAAAAGCTTG							
QY	623	gtctcagagatcaccacacacagtcgacagagatgctgcgcacacatagatattctctgc							
Db	283	GCTCCAGCAGTACACACACAGTCCAGAGAGTCCACACATCATATGATGCTCTGC							
QY	683	caaatccttcaggaagggcttgaatctccagctcctgcagctgaagaatttgcagcttc							
Db	343	CAATGCTCTTCAGAGAGGGGCTGAGATCTCCAGTCTGTCAGTGAAGATGATCAGTTC							
QY	743	gctgttgcgaacatgtgatttggaaagatttcaagaccgtctgtgtgcgcaactgcaaga							
Db	403	GCTGTTGCACACATCTGA-TTGGAGAGATTTCAGACCTGCTGTGTCGGCAGTCAAGCA							
QY	803	tgcgttgcgtgtgcgctgtgtgcacacctgcgtgcgtgaagaagtgttctatctcgaggacac							
Db	462	TGCGTTGGCTGTGCCTGCTGTGCACCTCTGCTGGAAGTGTCTATTCCGGGAGAC							
QY	863	catgccttcgagagctctgtctcgagatgggaaagatgcacacctctgatacatgaacg							
Db	522	CATGCCCTCGAGAGCTCTGTGTCGGATGGGGAAGATGCCACCTCTCTGATCATTAAGC							
QY	923	caacctggaagatgttggaaaggaagcagtggaagaacacacagacaacacgtccca							
Db	582	CACCTGGGAAGATGCTTGGAGAGAGAGACAGTGAAGAAAGACACACAGCAACAGTCCCA							
QY	983	agccatcagcgtgggagtg 1001							
Db	642	AGCCATCAGCGTGGGATG 660							

RESULT 9
 BG519751 884 bp mRNA EST 02-APR-2001
 LOCUS 602578902P1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
 DEFINITION mRNA sequence.
 ACCESSION BG519751
 VERSION BG519751.1 GI:13515513
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 884)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: L10M176 row: d column: 01
 High quality sequence stop: 859.
 Location/Qualifiers

FEATURES
 source
 1. 884
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3503184"
 /clone_lib="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dr priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 199 a 251 c 249 g 185 t

Query Match 37.1%; Score 371; DB 154; Length 884;
 Best Local Similarity 99.8%; Pred. No. 3.6e-186;
 Matches 491; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	176	attacttgcattctcgaagaattcattagttgagcgtcgcagcttcccaattccagcg							
Db	273	ATTACTTGGCAATCTTGAGAAATTCATAGTTGAGGCGCTCAGCTTCCAACTTCAGCAG							
QY	236	agcgtgcaggaatcagagagagtgctgcagagacgcgcagccccagacagaaagaat							
Db	333	AGCGTCCAGAGATACAGAGAGAGTGGCAGAGAGCGCCAGCCACAGCAAGAAAGT							
QY	296	cagtaaccagaatcatcttcccttggaaacaggtctgtccatcccgatgaagattcgaaat							
Db	393	CAGTACCCGAATAATCATCTTCTTGGAGACAGGCTGCCATCCGATGAAGATTGAAAT							
QY	356	gtcagtgccacactgttcaacataagccccagacagctctctgattatgacgtgtgtg							
Db	453	GTCAAGTCCACACTTGTCAACATTAAGCCCCGACACACTCTGTCTACTGTGAGTGTGAG							
QY	416	ggcacattggggcagctgtgcccgtcattcagagacagagtgtagagggctctgagacc							
Db	513	GGCACAATTGGGCGAGCTGTGGCTCATTTACGAGACAGAGTGGACAGGGTCTCGGACAC							
QY	476	ctgagctgtgtgttgtgtccacctgagcagcagatcaccaacagcgttgcgaagat							
Db	573	CTGGCTGT							
QY	536	cttgcgcagagagagacgcgcttgcgcatcttgggaaagcgcgttccaccttgcgtgt							
Db	632	CTTGCAG							
QY	596	ggttgcac							
Db	692	GGTTGCCCCCAACACAGCTCAAAAGCTGGCTCCAGCAGTACACACAGCAATGCCAGAGGT							
QY	656	cctgcac							
Db	752	CCGCGACACACAT 763							

RESULT 10
 BE900936 745 bp mRNA EST 29-SEP-2000
 LOCUS 601674206P1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3957240 5',
 DEFINITION mRNA sequence.
 ACCESSION BE900936
 VERSION BE900936.1 GI:10389609
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 745)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory

(MGC)

JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1307 row: P column: 16
High quality sequence stop: 736.
Location/Qualifiers

FEATURES
Source
1. .920
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583679"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 196 a 261 c 272 g 190 t 1 others
ORIGIN

Query Match 33.7%; Score 337; DB 153; Length 920;
Best Local Similarity 99.8%; Pred. No. 4.6e-168;
Matches 457; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 tcgcagcacaagatcaaacacagctcaactcaccacccgagacattccctcgt 60
DB 245 TCGCAGCCCAAGATTCAACCCAGCTCAACCTCAACCCGAGACATCTCCCTGCT 304
QY 61 caccagtttcgcgtctgaagaagagggcccccacccctcagtggtccatggtcaggtga 120
DB 305 CACCAAGTTCCCGCTGTAAAGAGAGAGGCCCCACCTCACTGTGCCCCATGTTACGGGTGA 364
QY 121 atgctcctcctcaagctacagctcgtcccaagagagagtgagcagagagatccattatc 180
DB 365 ATGCTCTCTCAAGTACACAGCTCCGTCCAGGAGGAGTGCGACAGAGGATGCCATTATTAC 424
QY 181 ttgcaatccttgaggaattcattagtgaggcgctgcagctcccaactccagcagaagct 240
DB 425 TTGCAATCCTTGAGGAATTCATTAGTTGAGCGCTGCGAGCTTCCCACTTCCACAGAGCGT 484
QY 241 gcaagagtaacaagagagtgagcgacagagcccgacagagagaaagaagtaagta 300
DB 485 GCAGGAGTACAGAGAGAGTGGCCAGAGCGCCAGGCCAGAGAGAAAGAAAGTACAGTA 544
QY 301 cccagaataatcattcctcttggaacagaggtctgcacatcccatgaagaattcgaatgtcag 360
DB 545 CCCAGAAATCATCTTCCCTTGGAAACAGGCTCG-CATCCCGATGAAGATTCGAATGTACG 603
QY 361 tgcacaacttgcaacataaagcccgacagcgtctctgtactgtgactgtgtgagagac 420
DB 604 TCCCACTACTTGTCAACATTAAGCCCCGACAGTCTGTCTACTGACTGTGAGGGCAGC 663
QY 421 atttgagcagctgtgcctcattacaggaacagagtg 458
DB 664 ATTGGGCGACTGTGCTGCTCATTTACGGAGACCAAGGTGG 701

RESULT 13
BE747163 992 bp mRNA EST 15-SEP-2000
LOCUS BE747163
DEFINITION 601577254F1 NIH_MGC_9 Homo sapiens CDNA clone IMAGE:3838237 5',

ACCESSION mRNA sequence.
BE747163
VERSION BE747163.1 GI:10161155
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 992)
NIH-MGC http://imgc.ncl.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DPF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM524 row: 1 column: 14
High quality sequence stop: 781.
Location/Qualifiers

FEATURES
Source
1. .992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3838237"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 225 a 287 c 313 g 167 t
ORIGIN

Query Match 33.0%; Score 330; DB 139; Length 992;
Best Local Similarity 99.6%; Pred. No. 2.5e-164;
Matches 500; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 tcgcagcacaagatcaaacacagctcaactcaccacccgagacattccctcgt 60
DB 148 TCGCAGCCCAAGATTCAACCCAGCTCAACCTCAACCCGAGACATCTTCCCTGCT 207
QY 61 caccagtttcgcgtctgaagaagagggcccccacccctcagtggtccatggtcaggtga 120
DB 208 CACCAAGTTCCCGCTGTAAAGAGAGAGGCCCCACCTCACTGTGCCCCATGTTACGGGTGA 267
QY 121 atgctcctcctcaagctacagctcgtcccaagagagtgagcagagagatgcatattac 180
DB 268 ATGCTCTCTCAAGTACACAGCTCCGTCCAGGAGAGTGGCAGAGGATGCCATTATTAC 327
QY 181 ttgcaatccttgaggaattcattagtgaggcgctgcagctcccaactccagcagaagct 240
DB 328 TTGCAATCCTTGAGGAATTCATTAGTTGAGCGCTGCGAGCTTCCCACTTCCACAGAGCT 387
QY 241 gcaagagtaacaagagagtgagcgacagagcccgacagagagaaagaagtaagta 300
DB 388 GCAGGAGTACAGAGAGAGTGGCCAGAGCGG-CACGCCAGAGAGAAAGAAAGTACAGTA 446
QY 301 cccagaataatcattcctcttggaacagaggtctgcacatcccatgaagaattcgaatgtcag 360
DB 447 CCCAGAAATCATCTTCCCTTGGAAACAGGCTCGCATCCCGATGAAGATTCGAATGTACG 506
QY 361 tgcacaacttgcaacataaagcccgacagcgtctctgtactgtgactgtgtgagagac 420
DB 507 TCCCACTACTTGTCAACATTAAGCCCCGACAGCTTGTCTACTGACTGTGAGGGCAGC 566

```

OY 421 attgggcagctgtgcgcattacagagaccaggtgagaggggtctctggaccctggc 480
    |||||||
Db 567 ATTTGGGACGCTGTGCGTATTCGAGACGAGTGAGACAGGGTCTTGGGCACCTGGC 626

OY 481 tgcgtgttgtgtccacctg 502
    |||||||
Db 627 TGCTGTGTGTGTCCACCTG 648

RESULT 14
LOCUS BE338283 878 bp mRNA EST 22-NOV-2000
DEFINITION 602035507F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4183488
VERSION 5', mRNA sequence.
ACCESSION BE338283
KEYWORDS BE338283.1 GI:11284683
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS 1 (bases 1 to 878)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9499 row: n column: 01
High quality sequence stop: 631.
Location/Qualifiers
1.878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4183488"
/clone.lib="NCI_CGAP_Brn64"
/tissue_type="glioblastoma with EGFR amplification"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.57 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 208 a 250 c 230 g 190 t

ORIGIN
Query Match 32.4%; Score 324; DB 147; Length 878;
Best Local Similarity 99.8%; Pred. No. 3.8e-161;
Matches 444; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 tgcgagccacaagatccaacccagctcaacctcaccacccggacatcttccccctgct 60
    |||||||
Db 175 TCGCAGCCACCAAGATTCACACCCAGCTCAACCTCATCCACCCGAGCATCTTCCCGCTGCT 234

OY 61 caccagtttcgctgttaagaagagggcccccacccctcagtggtccatggttcaaggtga 120
    |||||||
Db 235 CACCACATTCCTCCGTGAAGAGAGGGCCACCCTCAGTGTGCCCAATGGTTAGGGTGA 294

OY 121 atgccccctcaagtaccagctcgtctcccaagagaggtgacagagatgcatattac 180
    |||||||
Db 295 ATGCTCTCTTAATACCAAGTCCGCTCCCAAGAGAGGAGTGGCAGAGGATGCCATTATTAC 354

OY 181 ttgcaatcctgagaaattcattagttgagcgctgcagcttcccaactccaagagagcgt 240
    |||||||
Db 355 TTGCATATCTGAGGAATTCATAGTGAAGGCGGTGACAGCTTCCCAACTTCCAGCAGAGCGT 414

```

```

OY 241 gcagagatcagaggagtgctgcaggacggccagcccgacagagaaagaaatcagta 300
    |||||||
Db 415 GCAGGAGTACAGAGAGAGTCCGACAGACGG-CCAGCCCCGACGAGAGAAAGAGTACGTA 473

OY 301 cccagaatcatcttctcttggaaacagagtgctccatcccgatgaagaattcgaaatgtcag 360
    |||||||
Db 474 CCAGAAATCATCTCTCTTGGAAACAGGGTGTGCGCATCCGATGAAGATTGGAATGTGAC 533

OY 361 tgcacacactgtacaataagcccgacacagctctctgtactgtgactgtgtgagggcac 420
    |||||||
Db 534 TGCACACATCTGTCAACATAAGCCCGACAGCTGTGCTACTGTGAGTGTGTGAGGGCAC 593

OY 421 attgggcagctgtgcgcattac 445
    |||||||
Db 594 ATTTGGGACGCTGTGCGTATTCAC 618

RESULT 15
LOCUS BE250309 574 bp mRNA EST 13-JUL-2000
DEFINITION 600943455F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960077 5',
VERSION BE250309
KEYWORDS BE250309.1 GI:9120418
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS 1 (bases 1 to 574)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM52 row: n column: 14
High quality sequence stop: 571.
Location/Qualifiers
1.574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2960077"
/clone.lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Site-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 124 a 158 c 167 g 119 t 6 others

ORIGIN
Query Match 31.7%; Score 317; DB 165; Length 574;
Best Local Similarity 99.7%; Pred. No. 2e-157;
Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 503 gcaacgagatcaccacacagggcttgcagatattctgtcagagagaaagcgcttggc 562
    |||||||
Db 68 GCACGCGAGATCACACACAGGGGCTTGGCCAAAGTATCTTCTCAGAGAAACGCGCTTGGC 127

OY 563 atcttggaaagccggttacacctgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 622
    |||||||

```

Db 128 ATCTTTGGGAAAGCCGCTTACCCCTTTGCTGCTGTTGCTGCCCCCAACCAAGCTCAAAAGCCTG 187
Oy 623 gctccagagctaccacacacagctgccaagaggttccctgcacacacatgaatattctctc 682
Db 188 GCTCCAGCAGTACACACAGTGCAGAGAGGTCTTGCACCAATCATGATGATCTCTGC 247
Oy 683 caatgccttcaggaagggtctgagatctccagctcctgcagtggaaagatltacagtic 742
Db 248 CAAATGCTTTAGGAAGGGGCTGAGATCTCCAGTCTGCAGTGGAAAGATGTGATCAGTTTC 307
Oy 743 gctgttggaacaatctgatttggaaagatttcaagacctctcgtgctgagcaactgcaagca 802
Db 308 GCTGTGGCCAAAGCTGTGATTTGGAAAGATTTCAGACCTGCTGTGCTGGGCGACATGCAAGCA 367
Oy 803 tgcgtttgctgtgcgctgtgacacacctctgcgctggaagtggtctattccggggagac 862
Db 368 TCGGTTTGCTGTGCTGCTGTGCACACCTCTGCTGGAAGTGTCTATTCCGGGAGAC 427
Oy 863 catgcct 870
Db 428 CATGCCCT 435

RESULT 16
AUI24671 762 bp mRNA EST 23-OCT-2000
LOCUS AUI24671 NT2RM4 Homo sapiens cDNA clone NT2M400035 5', mRNA
DEFINITION sequence.
ACCESSION AUI24671
VERSION AUI24671.1 GI:10949387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 762)
AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
,Y., Sugano,S., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3851
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Source
1. 762
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM400035"
/clone_1lb="NT2RM4"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"

BASE COUNT 180 a 212 c 202 g 165 t 3 others
ORIGIN

Query Match 31.3%; Score 313; DB 107; Length 762;
Best Local Similarity 99.7%; Pred. No. 2,7e-155;
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 tctgcagcacaaatccaaccagctcaacctcatccaccgagacatcttccctgtct 60

Db 289 TCGAGCCACAGATTTCAAAACCAAGCTCAACCTCATCCACCGAGACATCTTCCCTGCT 348
Oy 61 caccagcttcgctgttaagaagaggagggccacacctcagctgtgccaatggttcagggtga 120
Db 349 CACCAAGTTTCCCTGTAAAGAGAGGGCCCAACCTCATGTCCTCATGTTCCAGGGTGA 408
Oy 121 atgccttcctcaagttaccagctccgtccccaaggaggatgtgcagaggaatgccattac 180
Db 409 ATGCTCTCTCAAGTACAGCTCCGTCCTCAGAGAGAGTGGCAGAGAGATGCCATTATTAC 468
Oy 181 ttgcacatccttgaggaattcatagttgagcgctgtgagcttcccaacttccagcagaagcgt 240
Db 469 TTGCAATCTTGAGGAATTCATAGTTGAGGCGCTGCAGGCTTCCCAACTTCATCAGAGCGT 528
Oy 241 gcaagagatcagaagagagtgctgcagagagcccaagcccaagagaagaaagatcagta 300
Db 529 GCAGAGATACAGAGAGAGGCGCAGAGAGCGCCACCCACGACGAGAAAGATCAGTA 588
Oy 301 cccagaatcatcttcctctgacacaggtctgcacatcccatggaagattcgaaatcgac 360
Db 589 CCCAGAAATCATCTTCTTGGAACAGGCTCTCCATCCGATGAAGATTCCGAATGTACG 648
Oy 361 tgcg 364
Db 649 TGCC 652

RESULT 17
BF315158 1497 bp mRNA EST 21-NOV-2000
LOCUS BF315158 601899518P1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4128701 5',
DEFINITION mRNA sequence.
ACCESSION BF315158
VERSION BF315158.1 GI:11263379
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1497)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE NIH-MGC Arrayed by: The I.M.A.G.E. Consortium (ILNL)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@rs-femail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at: image.llnl.gov
Plate: LLCMI022 row: c column: 06
High quality sequence stop: 660.
FEATURES
Source
1. 1497
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4128701"
/clone_1lb="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

BASE COUNT 512 a 382 c 430 g 173 t
ORIGIN

Query Match 31.2%; Score 312; DB 147; Length 1497;
 Best Local Similarity 99.8%; Pred. No. 9.9e-155;
 Matches 432; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

43 gagacatccctccctcgtccacagattccgtctgtaagaaggagggcccccacccctcagtgt 102
 111 GGCACATCTTCCCTCCCTCCACGATTTCGGCTGTAGAGAGAGGGCCCACTCAGTGT 170
 Db 103 gccacatggttcaggtgtaagtctcctccaagtaaccagctccagagagaggtggca 162
 171 GCCCATGTTCAGAGGTGATGCTCTCAAGTACAGCTCCGTCACAGAGAGGAGTGGCA 230
 Oy 163 gaagagtcacattacttctgcaatctcgaagattcattagttgagagcgttgagcttcc 222
 231 GAGGATGCCATTATTACTTTCGATCTTGAGATTCTATTGTTGAGCGCTTGAGCTTCC 250
 Db 223 caactccagcagaagcgttcagagagtaacaggaagtgccagagagcccaagcccaagc 282
 291 CAACCTTCAGCAGAGCGTGCAGAGTACAGAGAGTGGCCAGAGCGG-CCAGCCCCAGC 349
 Oy 283 agagaaagaagtcagtaaccagaatacattctccttggaagaggtctgcatcccgat 342
 350 AGAGAAAGAAAGTCACTACCCAGAAATCATCTTCTTGGAACAGAGGTCTGCCATCCCGAT 409
 Db 343 gaagatgcgaatgtcagtgccacactgtcacaataagcccgagagcagctctgtact 402
 410 GAAGATTCGAATGTCAAGTCCACACTGTGCAACATTAAGCCCGACACGCTCTGCTACT 469
 Oy 403 ggaactgtgtgagagacatttgagcagctgtgcgtcattacagagagcaggttgagcag 462
 470 GGACTGTGTGAGAGGACATTTGGGACAGCTGTCCGTCATTAGGAGACACAGGTGAGCAG 529
 Db 463 ggtcctggggcacc 475
 530 GGTCTGTGGGCACC 542

RESULT 18
 BE902696 735 bp mRNA EST 29-SEP-2000
 LOCUS 601677393F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959926 5',
 DEFINITION mRNA sequence.
 BE902696
 ACCESSION BE902696.1 GI:10393148
 VERSION EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 735)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: Image.lnl.gov
 Plate: LNC841 row: 3 column: 23
 High quality sequence stop: 732.
 Location/Qualifiers
 1. 735

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3959926"
 /clone_lib="NIH_MGC_21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10b (phage-resistant)"
 /note="Organ: placenta; Vector: pORF7; Site_1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the Laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 168 a 206 c 203 g 157 t 1 others

Query Match 30.7%; Score 307; DB 141; Length 735;
 Best Local Similarity 100.0%; Pred. No. 4.3e-152;
 Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 183 gcaactcctgaggaattcattagttgagagcgtcagctcccaactccagcagaagctgc 242
 Db 134 GCAATCTTGAGGAATTCATGTGAGCGCTGCAGCTTCCCAACTTCCACAGAGCGTGC 193
 Oy 243 aggaatcacagagagagtgccagagagcccaagcagagagaaagaagtcagttacc 302
 194 AGAGATACAGAGAGAGTGCAGAGACGCGCCAGCCCGCAGAGAAAGATTCAGTACC 253
 Db 303 cagaatcatctccttgagacagaggtctgcatacccgatgaagattcgaaatgcatg 362
 254 CAGAAATCATCTTCTTGGAACAGAGGTCTGCATCCGATGAAGATTCGAATGTCACTG 313
 Oy 363 ccacactgtcaacataagcccgagacagctctcgtactggaactgtgtgagagacat 422
 314 CCACACTTGTCAACATTAAGCCCGACACGCTCTGCTGACTGAGTGTGAGAGGCGCAT 373
 Oy 423 ttggagcagctgtcgtcattacagagacaggttgagacagaggtcctggagacctgctg 482
 374 TTGGGACGCTGTCCGTCATTACGAGACCAAGTGAGACAGGCTCGGACCCCTGGCTG 433
 Db 483 ctgtgtt 489
 434 CTGTGTT 440

RESULT 19
 AA310236 431 bp mRNA EST 19-APR-1997
 LOCUS AA310236 Jurkat T-cells V Homo sapiens cDNA 5' end, mRNA sequence.
 DEFINITION EST181085
 AA310236
 ACCESSION AA310236.1 GI:1962584
 VERSION EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 431)
 AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
 C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Goodbye,J.D., White
 O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Well,C., Clayton,R.A.,
 Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,
 L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodok,A.,
 Guelim,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
 Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
 Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
 Dimarzio,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A., He,W.W.,
 Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
 Kunsch,C., Hungjun,J., Li,H., Weisner,P.S., Olsen,H., Raymond,L.,
 Mei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
 M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
 Venter,J.C.
 TITLE Initial assessment of human gene diversity and expression patterns
 JOURNAL based upon 83 million nucleotides of cDNA sequence
 MEDLINE Nature 377 (6547 Suppl), 3-174 (1995)
 96026280

DEFINITION 602015077F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:41507/1

5', mRNA sequence.
ACCESSION BF342802
VERSION BF342802.1 GI:11289829
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1124)
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaab-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1M9414 row: J column: 20
High quality sequence stop: 670.
Location/Qualifiers
1. 1124
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4150771"
/clone_lib="NCI-CGAP_Brn64"
/tissue_type="g1oblastoma with EGFR amplification"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pCMV-SPORE6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.57 kb. Constructed by Life
Technologies. Note: this is a NCI-CGAP Library."
BASE COUNT 322 a 281 c 346 g 175 t
ORIGIN
Query Match 27.9%; Score 279; DB 147; Length 1124;
Best Local Similarity 100.0%; Pred. No. 3.6e-137;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 625 tccaagcagaccacacacagctgcagagagctcgcacacacacatgatcctgcgca 684
DB 6 tccagacagraccaacacacagctgcagagagctcgcacacacacatgatcctgcgca 65
QY 685 aatgcctcagaagaagagctgagatcctcagctcagctgagaaagatgatcagctgcg 744
DB 66 AATGCTTCAGGAAGGGGCTGAGATCTCCAGTCTGCAAGTGAAGATTTGATCAGTTCCG 125
QY 745 tcttgagacaatgatcttggaaagattcagaactgtctgtgctgagcaatgatcag 804
DB 126 TGTTCGGAACATGTGATTGGAGAGTTTCAGACTGTCTGTGCGGCACTGCACATG 185
QY 805 cgttgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 864
DB 186 CGTTTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 245
QY 865 tgcctcagaagcctcgtgctcgaatgaggaagaatgcga 903
DB 246 TGCCCTCGAGGCTGTGCTGCGATGGGGAAGATGCCA 284
RESULT 24
LOCUS BE742908 677 bp mRNA EST 15-SEP-2000
DEFINITION 601574609F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3835658 5',
mRNA sequence.
ACCESSION BE742908
VERSION BE742908.1 GI:10156900
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 677)
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaab-r@mail.nih.gov
Tissue Procurement: DCTD/DPF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM518 row: a column: 03
High quality sequence stop: 672.
Location/Qualifiers
1. 677
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3835658"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 160 a 173 c 217 g 127 t
ORIGIN
Query Match 27.1%; Score 271; DB 139; Length 677;
Best Local Similarity 100.0%; Pred. No. 6.3e-133;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 669 agtatgattcctcgtccaatgcttcaggaagagctgagatcctcagctcagtgagaa 728
DB 1 AAGTATATTCCTCCCAATAGCTTCAGGAAGGGGCTGAGATCTCCAGTCTCGAGTGGAA 60
QY 729 agattgataagctgctgttggaacaatgatttggaaagatttcaagaccgtctgtg 788
DB 61 AGATTGATCAGTTCGCTGTGCGAATCATGTGATTGGAAGAGTTTCAGACTGTCTGTG 120
QY 789 cggcactgcaagaatcgcttgctgtgctgtgctgtgctgtgctgtgctgtgctgtgct 848
DB 121 CGGCACTGCAGACATCGCTTTGGCTGTGCGTGCTGTGCTGTGCTGTGCTGTGCTGTG 180
QY 849 taltccggggaacacatgctcctgagagctcgtgctcgaatgaggaagaatgcacacctc 908
DB 181 TATTCGGGGAGACACATGCGCTCGAGGCTCTGCTCGGATGGGGAAGATGCACACCTC 240
QY 909 ctgatacatgaagcaacccttggaagatggtt 939
DB 241 CTGATACATGAAGCCACCTGGAAGATGTT 271
RESULT 25
LOCUS R55841 482 bp mRNA EST 23-MAY-1995
DEFINITION y989d01.r1 Soares infant brain IN1B Homo sapiens cDNA clone
IMAGE:40931 5' similar to SP:YK59.YEAST P36159 HYPOTHETICAL 96.8 KD
PROTEIN IN SIS2-MTD1 INTERGENIC ;, mRNA sequence.
ACCESSION R55841
VERSION R55841.1 GI:825947
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 482)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucada, T., Le, M., Lennon, G., Maira, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Insert Size: 1777
 High quality sequence stops: 387 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1777 Std Error: 0.00
 Seq primer: M13Rpl
 High quality sequence stop: 387.
 Location/Qualifiers
 1. 482
 /organism="Homo sapiens"
 /db_xref="GDB:413472"
 /db_xref="taxon:9606"
 /clone="IMAGE:40931"
 /clone_lib="Soares Infant brain INIB"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: lacMid BA; Site: 1: Not I; Site: 2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' ACTGTGAGAAATTCGCGCGCCGACAGAAATTTTCTTTTCTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the lacMid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 107 a 139 c 131 g 102 t 3 others
 ORIGIN

Query Match 26.2%; Score 262; DB 187; Length 482;
 Best Local Similarity 100.0%; Pred. No. 3.8e-128;
 Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 gcaagagtagcaggaggtgacgagagcgcacccacacagagaaagaatcaata 300
 |||||||
 DB 43 GCAGGAGTACGAGGAGGTGCGAGGAGCGCCACCCACAGAGAAAGTCAAGTA 102

QY 301 cccagaaatcatcttcctctggaacaggtctgcccattcccgatgaagattcgaaatgtaag 360
 |||||||
 DB 103 CCCAGAAATCATCTTCTTGGACAGAGGTCTGCCATCCCGATGAAGATTCCAAAATGTAG 162

QY 361 ttgcacactgtcaacataagcccccagacagtcctgtactggaactgtgtgtagggac 420
 |||||||
 DB 163 TGCACACTGTCAACATAAGCCCCGACACGCTCTCTACTAGCATGTGTGAGAGGCAC 222

QY 421 attgggagagctgtgcccgtactacggagacaggtgacaggggtccctgggacacctggc 480
 |||||||
 DB 223 ATTTGGGAGGTGTGCGCTCATTTACGGAGACAGGTGAGGTCTGGGACCCCTGGC 282

QY 481 tgcctgtctgtgtccacactg 502
 |||||||
 DB 283 TGCCTGTCTTGTGTGCCACCTG 304

RESULT 26
 BF969043
 LOCUS BF969043 944 bp mRNA EST 22-JAN-2001

DEFINITION 60226966F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4357985 5', mRNA sequence.
 ACCESSION BF969043
 VERSION BF969043.1 GI:12336258
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 944)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LTM995 row: d column: 18
 High quality sequence stop: 591.
 Location/Qualifiers
 1. 944
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4357985"
 /clone_lib="NIH_MGC_84"
 /tissue_type="adrenal cortex carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 200 a 266 c 264 g 214 t
 ORIGIN

Query Match 25.3%; Score 253; DB 172; Length 944;
 Best Local Similarity 100.0%; Pred. No. 2.5e-123;
 Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcagcacaagattcaaacccagctcaacctcaaccacccggacatcttccctgct 60
 |||||||
 DB 345 TCGCAGCCACAAGATTCAAAACCCACCTCATCCACCCGACATCTTCCCTGCT 404

QY 61 caccagtttcgctgtaagaagagagggcccccctcagtggtgccatggttcaaggtga 120
 |||||||
 DB 405 CACCACTTCCGCTCTTAAGAAAGAGGAGGCCCCCTCAGTGTGCCATGTTCAGAGTGA 464

QY 121 atgctctcctaagtcacagctcgcgtcccaagagaggtgagcagaggaatgcatattac 180
 |||||||
 DB 465 ATGCTCTCTCAAGTACACAGCTCCGTTCCAGAGGAGAGTGGCAGAGGATGCCATTATTAC 524

QY 181 ttgcaatcctgaagaattcaatagttgagggcgtcagcttcccaatccagcagagcgt 240
 |||||||
 DB 525 TTGCAATCCTGAGGAATTCATGATTGAGAGCGCTGACAGTCCCAACTCCACAGAGCGT 584

QY 241 gcaagagtagcagg 253
 |||||||
 DB 585 GCAGGAGTACAGG 597

RESULT 27
 A1991599/c
 LOCUS A1991599 517 bp mRNA EST 08-MAR-2000
 DEFINITION ws18c04.x1 NCI_CGAP_G6 Homo sapiens cDNA clone IMAGE:2497542 3', similar to SW:YATA_SCHPO 010155 HYPOTHETICAL 90.6 KD PROTEIN
 C1D4.10 IN CHROMOSOME 1. ;, mRNA sequence.
 ACCESSION A1991599

VERSION AI991599.1 GI:5838504
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 517)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapdb-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/db/brp/image/image.html
Insert length: 625 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 426.
Location/Qualifiers
1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2497542"
/clone_lib="NCI-CGAP_GC6"
/issue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pF7P3D-Pac (Pharmacia) with a modified
polylinker. Plasmid DNA from the normalized library
NCI-CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (clonoids 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 118 a 139 c 123 g 136 t 1 others
ORIGIN

Query Match 24.7%; Score 247; DB 104; Length 517;
Best Local Similarity 100.0%; Pred. No. 3.7e-120;
Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 658 tgcaccacatcaatgatgatcctcgtcccaatgcttcaagaaaggcctgaatctcagtc 717
DB 447 TGCACCACTAGATGATTCCTGCGCAAAATGCTTCACGAAGGGGCTGAGATCTCCAGTC 388
QY 718 ctgcagtggaagaatgatcatgctcgtgttcgaaacatgtgatttggaaagatttcaaga 777
DB 387 CTGCACTGGAAGAATGATGATCAGTTCCGTGTGGCAACATGTGATTGGAAGATTTCAGA 328
QY 778 ccgtgttgtgtggcactgaagcagatgctgtgtgctgtgtgctgtgtgcaacctctgct 837
DB 327 CCTGTGTGTGGCGGCGACTGCAACGATGCGTTGGCTGTGGCGTGTGCACACCTCTGGCT 268
QY 838 ggaagaatgtctatctccggggagacacatgacctgagaggtctgtgtccggatgggaaag 897
DB 267 GGAAGATGTCTATTCTCCGGGGACACCATGCCCTGCGAGGCTCTGTCCGATGGGAAAG 208
QY 898 atggcac 904
DB 207 ATGCCAC 201

RESULT 28
AA634909

LOCUS AA634909 446 bp mRNA EST 21-OCT-1997
DEFINITION ab27h02.r1 Stratagene lung (#937210) Homo sapiens cDNA clone
IMAGE:842067 5' similar to SW:YK59.YEAST P36159 HYPOTHEETICAL 96.8
KD PROTEIN IN SIS2-MTD1 INTERGENIC REGION. ; mRNA sequence.
ACCESSION AA634909
VERSION AA634909.1 GI:2558123
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 446)
Hiller, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilison RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 430.
Location/Qualifiers
1..446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:842067"
/clone_lib="Stratagene lung (#937210)"
/sex="male"
/dev_stage="72 years"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT, normal lung. Average insert size: 1.0 kb; Uni-7AP XR
Vector; ~5' adaptor sequence: 5' GAATTCGACGACG 3' ~3'
adaptor sequence: 5' CTCGAGTTTATTTTATTTT 3'."

BASE COUNT 105 a 112 c 131 g 97 t 1 others
ORIGIN

Query Match 24.6%; Score 246; DB 10; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.3e-119;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 756 tgtgatttgaagaagtacagacctgtctgtgtgcgcacatgacaagcatgcttggctgt 815
DB 59 TGTGATTTTGAAGAGTTTTCAGACCTGTCTGTGTCGGCAGCTGCAAGCATGCTTTGGCTGT 118
QY 816 ggcgtgtgtcacacacctctgtgctgaaagtgtctatctccggggagacacatgacctgcgag 875
DB 119 GCGCTGTGTGACACCTCTGTGCTGGAAGTGTCTATTCCGGGGACACCAATGCCCTCGAG 178
QY 876 gctctgtgtccgagtggaagaatgccaacctctctatatatgaagccacctggagaat 935
DB 179 GCTCTGTGTCCGATGAGGAAAGATGCGACCTCTCTATACATGAACCAACCCCTGGAAGAT 238
QY 936 ggtttgaaagaaagcagtgtgaaagaagacacagacacaaagctcccaagccatcagcgtg 995
DB 233 GGTTTGAAGAGAGACAGTGTGAAAAAGACACAGCACAAACGTTCCCAAGCATACAGCTG 298
QY 996 gggatg 1001
DB 299 GGCATG 304

RESULT 29
AA634909

LOCUS AM889463 346 bp mRNA EST 24-MAY-2000
 DEFINITION RC6-NT0029-240400-011-B07 NT0029 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM889463
 VERSION AM889463.1 GI:8053668
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 346)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 JOURNAL MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC6-NT0029-240400-011-B07&ts=2000-04-24&ft=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 12
 High quality sequence stop: 310.
 FEATURES
 source
 1..346
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT0029"
 /dev_stage="Adult"
 /note="Organ: nervous tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 75 a 103 c 85 g 83 t
 ORIGIN
 Query Match 24.3%; Score 243; DB 122; Length 346;
 Best Local Similarity 100.0%; Pred. No. 4.8e-118;
 Matches 243; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 641 ccagtcgcaagagtcctgcacacacatgatgatcttcgccaatgccttcgaagaag 700
 |||
 |||
 Db 335 ccagtcgcaagagtcctgcacacacatgatgatcttcgccaatgccttcgaagaag 276
 Oy 701 ggcctgagatcctcagctcctcagtggaagaatgatcagttgctgttggaacatgta 760
 |||
 |||
 Db 275 ggcctgagatcctcagctcctcagtggaagaatgatcagttgctgttggaacatgta 216
 Oy 761 ttggaagaattcagaagcctctgtgtgcgacactgcaagcatgcttgctgtgcct 820
 |||
 |||
 Db 215 ttggaagaattcagaagcctctgtgtgcgacactgcaagcatgcttgctgtgcct 156
 Oy 821 ggtgcacacctgtgctgtaagtgctatctccggggaacacatgcctgtgcgactct 880
 |||
 |||
 Db 155 ggtagcacacctgtgcgtaagtgctatctccggggaacacatgcctgtgcgactct 96
 Oy 881 ggt 883

Db 95 GGT 93
 |||
 RESULT 30
 AA311855
 LOCUS AA311855
 DEFINITION EST182568 Jurkat T-cells VI Homo sapiens cDNA 5' end, mRNA sequence.
 ACCESSION AA311855
 VERSION AA311855.1 GI:1964184
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 501)
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghegan,N.S., Glodde,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utechtack,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Meli,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
 Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl), 3-174 (1995)
 JOURNAL MEDLINE 96026280
 COMMENT Other_ESTS: TH0175624
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/ngi/ngi.html>)
 Seq primer: M13 Reverse.
 FEATURES
 source
 1..501
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):159063"
 /db_xref="taxon:9606"
 /clone_lib="Jurkat T-cells VI"
 /cell_type="T-lymphocyte"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 125 a 148 c 125 g 98 t 5 others
 ORIGIN
 Query Match 24.0%; Score 240; DB 5; Length 501;
 Best Local Similarity 100.0%; Pred. No. 2e-116;
 Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 tcgcagccacaagaattcaaacccagctcaacatcatcaccgggaacatctcccccgtct 60
 |||
 |||
 Db 110 tcgcagccacaagaattcaaacccagctcaacatcatcaccgggaacatctcccccgtct 169
 Oy 61 caccagttcgctgtgaagaggaggccaccactcagttgcccacatggtcagggtga 120
 |||
 |||

```
Db 170 CACCAGTTCCGCTGTAGAGAGAGGAGGCCACCCCTCAGTGTGCCANTGTTCCAGGCTGA 229
Oy 121 atgctctcctcaagtagcagctccctcccaaggagtagcgagggatgcatattaac 180
    |||
Db 230 ATGCTCTCCCAAGTACCACTCCGTCCAGAGGAGTGGCAGAGGAGGATGCCATTATTATAC 289
Oy 181 ttgcaatccttgaggaattatagttgagcgctgtagctcccaactcccaagcagcgt 240
    |||
Db 290 TTGCAATCTCTGAGGAATTCATAGTTGAGGCGCTGCACGCTTCCCACTTCCACAGAGCGCT 349

RESULT 31
BE795434 698 bp mRNA EST 20-SEP-2000
LOCUS 601592991P1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3946774 5',
DEFINITION mRNA sequence.
ACCESSION BE795434
VERSION BE795434.1 GI:10216632
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at: Image.lnl.gov
Plate: L1CM807 row: f column: 23
High quality sequence stop: 389.
Location/Qualifiers
1..698
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3946774"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="pDH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```
BASE COUNT 201 a 212 c 175 g 110 t
ORIGIN
```

```
Query Match 23.0%; Score 230; DB 140; Length 698;
Best Local Similarity 100.0%; Pred. No. 4.3e-111;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 503 gacgacacatcaccaacagggcttcgcaagtagctgtcgagagagaaacggccttgc 562
    |||
Db 163 GCACGCAAGTACCAACACGGGCTTGCCAAAGTATCTTGCTGCAGAGAGAAAGCGCCTTGCC 222
    |||
Oy 563 atctttggaagaagccgcttcaccccttctgtgtgttgcacccaacacagctcaagcctg 622
    |||
Db 223 ATCTTTGGGAAGCGGCTTCAACCTTTGCTGTGTGTGTCGCCCAACACAGCTCAAGAGCTG 282
    |||
Oy 623 gctcagcagtagtaccacacagtagtccagagagtagtctgacacacatcatgattcttc 682
    |||
Db 283 GCTCCAGCAGTACCAACACAGTGCAGAGAGGTCTCTGCACCAATCATCATGATTCTCTGC 342
```

```
Oy 683 caatgctcctcaggaaggctgagatctccagctcctcgtcagtggaagat 732
    |||
Db 343 CAAATGCTTTCAGGAAGAGGCGCTGAGATCTCCAGTCTCTGAGTGGAAAGAT 392
```

```
RESULT 32
BE260626 938 bp mRNA EST 26-OCT-2000
LOCUS 601146116P1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3161691 5',
DEFINITION mRNA sequence.
ACCESSION BE260626
VERSION BE260626.1 GI:9132065
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov
Plate: L1CM119 row: o column: 04
High quality sequence stop: 621.
Location/Qualifiers
```

```
1..938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3161691"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="pDH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
```

```
BASE COUNT 190 a 263 c 283 g 202 t
ORIGIN
```

```
Query Match 22.3%; Score 223; DB 165; Length 938;
Best Local Similarity 100.0%; Pred. No. 2.4e-107;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 tggcagccacacagattcaaacccagctcaactcatccacccggagacattcccccgt 60
    |||
Db 359 TCGCAGCCCAAGATTCANACCCAGCTCAACTCATCCACCCGGACATTTCCCCCTGCT 418
    |||
Oy 61 caccagttccgctgtaaagaaggagggccacccctcagttgtgccatgttcagggtga 120
    |||
Db 419 CACCAAGTTCCGCTGTAAAGAAAGAGGCGCCCAACCGCTCAAGTGTGCCAATGTTCCAGGCTGA 478
    |||
Oy 121 atgctcctcctcaagtagcagctcgctcccaaggagagtagtgcagagagtagtccattat 180
    |||
Db 479 ATGCTCTCCCAAGTACCAAGCTCCGTCCAGAGGAGTGGCAAGAGGAGTGCATTATTATAC 538
    |||
Oy 181 ttgcaatccttgaggaattatagttgagggcggtgagcttccc 223
    |||
Db 539 TTGCAATCTCTGAGGAATTCATAGTTGAGGCGCTGCAAGCTTCCC 581
```



```

RESULT 33
AM806551 278 bp mRNA EST 17-MAY-2000
LOCUS IL0-ST0002-160599-003 ST0002 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM806551
ACCESSION AM806551.1 GI:7899550
VERSION EST.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 278)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brijnes,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=IL0-ST0002-160
599-003&l3=1999-05-16&l4=1)
Seq primer: puc 18 forward
High quality sequence stop: 278.
FEATURES
source
1..278
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="ST0002"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; site_1: SmaI;
site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196/716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 64 a 84 c 70 g 60 t
ORIGIN
Query Match 20.9%; Score 209; DB 120; Length 278;
Best Local Similarity 100.0%; Pred. No. 6.1e-100;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcgcagccacaagaatccaacccagctcaacatcatcacccggagacatctccctgct 60
|||||
DB 70 TCGGAGCCACAAGATCAAAACCCAGCTCAACCTCATCCACCCGGAGCATCTCCCTCT 129
|||||
QY 61 caccagttccgcgtgtaagaaggagcccaacctcaagtggtgccatggttcaggatga 120
|||||
DB 130 CACACAGTTTCCGCTGTAAAGAGGAGGCCCCACCTCAGTGTGCCCAAGGTTACAGGGTGA 189
|||||
QY 121 atgcctctcaagaagcagctccgctcccaaggaggaftgagagaaggagatgcattatcac 180
|||||
DB 190 ATGCTCTCTCAAGTACCAAGCTCCCTCCAGAGGAGAGTGGCAGAGGATGCCATTATTATAC 249
|||||
QY 181 ttgcaatcctgaggagaatcatagttgag 209
|||||
DB 250 TTGCAATCCTGAGGAATTCATAGTTGAG 278

```

```

RESULT 34
T34024 282 bp mRNA EST 06-SEP-1995
LOCUS EST61387 Human white blood cells Homo sapiens cDNA 5' end similar
DEFINITION T34024
ACCESSION T34024
VERSION T34024.1 GI:616122
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 282)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W., Clayton,R.A.,
Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.M., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinklejr,P.S., Kelley,J.M.,
Klimek,K.W., Kelley,J.C., Liu,L.-T., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Uteback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J.,
Dimke,D., Feng,P., Ferrie,A., Fischer,C., Hastings,G.A., He-W.-W.,
Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Korak,D.L.,
Kunsch,C., Li,H., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei
,Y.-F., Wang,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
Initial Assessment of Human Gene Diversity and Expression Patterns
Based Upon 83 Million Basepairs of cDNA Sequence
JOURNAL Nature 377, 3-174 (1995)
MEDLINE 96026280
COMMENT Other ESTs: TFC15444
Contact: Venter, JC
The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@tdb.tigr.org
For clone availability, additional sequence and expression
information related to this EST, please contact the TIGR Database
(tdbinfo@tdb.tigr.org)
Seq primer: M13 Reverse.
FEATURES
source
1..282
location/Qualifiers
/organism="Homo sapiens"
/db_xref="ATCC (host):104694"
/db_xref="taxon:9606"
/clone_id="Human white blood cells"
/tissue_type="white blood cells"
/note="Organ: blood"
BASE COUNT 62 a 83 c 73 g 62 t 2 others
ORIGIN
Query Match 20.9%; Score 209; DB 188; Length 282;
Best Local Similarity 100.0%; Pred. No. 6.1e-100;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 503 gacagcagataccacacacggtctgcaagatattctgtgcagagagaagcgcttggc 562
|||||
DB 45 GCACGCAATGATCACACAGCGGCTCCCAAGTATCTTCTGCGAGAGAAACGCGCTTGGC 104
|||||
QY 563 atcttgggaagcgcgttcaccccttctgctggtgttcccccaacgcgtcaagcctg 622
|||||
DB 105 ATCTTTGGAAAGCCGCTTACCCCTTGTGCTGTGTGTTTCCCCCAACGACGAAAGCCCTG 164
|||||
QY 623 gctccagagatcacacacacagtgccagagagtgcttcgacacacatcatatctctgc 682
|||||
DB 165 GCTCCAGCAGTACCAACCAACGAGTGCAGAGGTCTCTGCACCAACATCAGATGATTCCTGC 224

```

QY 683 caaatgccttcaggaaaggcgctgagatct 711
 |||
 Db 225 CAAATGCCTTCAGGAAGGGCTGAGATCT 253

RESULT 35
 BG480926
 LOCUS BG480926 963 bp mRNA EST 21-MAR-2001
 DEFINITION 602530075F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4653330 5',
 mRNA sequence.
 ACCESSION BG480926
 VERSION BG480926.1 GI:13413205
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 963)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rrmail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: L1CMI439 row: f column: 19
 High quality sequence stop: 658.
 Location/Qualifiers
 1..963
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4653330"
 /clone_lib="NIH_MGC_21"
 /tissue_type="Choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pORF7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 270 a 246 c 291 g 155 t 1 others
 ORIGIN

Query Match 20.6%; Score 206; DB 154; Length 963;
 Best Local Similarity 100.0%; Pred. No. 2.7e-98;
 Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 734 gatcgccttcgttcgacacatgtattggaaggttcagacctgtcgtgcgca 793
 |||
 Db 2 GATCAGTTCGCTTGCGCAACATGATTTGGAAGAGTTTCAGACCTGCTGTGGCGCA 61

QY 794 ctgcaagcatcgcttgcgtcgtcgttcgacaccttgcgtggaagtgtctatc 853
 |||
 Db 62 CTGCAAGCATGCGTTGGCTGTGCTGTGCAACACCTCTGGCTGGAAGGTGATATTC 121

QY 854 cggggaaccatcgccctcgagagctcgttcgagatggggaagatgccacctctcat 913
 |||
 Db 122 CGGGGACACACATGCTCGAGGCTCTGCTCGGATGGGGAAGATGCCACCTCTCTGAT 181

QY 914 acatgaagccacctggaagatgatt 939
 |||
 Db 182 ACATGAAGCCCTTGGAAGATGATT 207

RESULT 36
 R51138
 LOCUS R51138 472 bp mRNA EST 18-MAY-1995
 DEFINITION Y971c08.r1 Soares Infant Brain INTB Homo sapiens cDNA clone
 IMAGE:38752 5' similar to SP:YK59_YEAST P36159 HYPOTHETICAL 96.8 KD
 PROTEIN IN SIS2-MTD1 INTERGENIC ;, mRNA sequence.
 ACCESSION R51138
 VERSION R51138.1 GI:813040
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 472)
 Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
 M., Hultman,M., Kucba,T., Le,M., Lennon,G., Morris,M., Parsons,J.,
 Rifkin,J., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston
 R., Williamson,A., Wohldmann,P. and Wilson,R.
 The WashU-Merck EST Project
 Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 High quality sequence stop: 327
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: M13RP1
 High quality sequence stop: 327.
 Location/Qualifiers
 1..472
 /organism="Homo sapiens"
 /db_xref="GDB:411293"
 /db_xref="taxon:9606"
 /clone="IMAGE:38752"
 /clone_lib="Soares Infant Brain INTB"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: latmid BA; Site_1: Not
 I; Site_2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer [5];
 AACGGAAGATTCGCGCCGACGAGATTTTATTTTATTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (Pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the latmid BA vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 105 a 131 c 125 g 102 t 9 others
 ORIGIN

Query Match 19.9%; Score 199; DB 187; Length 472;
 Best Local Similarity 100.0%; Pred. No. 1.3e-94;
 Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 gcaagagtaacagagagtgacgacagagccacgacagagaagaagatcagta 300
 |||
 Db 43 GCAGGAGTACAGAGAGAGGCGCAGAGCGCCACGACGAGAAAGAGTCAGTA 102

QY 301 cccagaatcatcttccttcggaacagagtgctccatccgatgaagatcgaatgcag 360
 |||
 Db 103 CCCAGAAATCATCTTCTTGGAACAAGGCTCCATCCCGATGAAGATTGAAATGTCAG 162

QY 361 tggcacactgtcaacaataaagcccgacacgctctgtcactcgaagcttggtagggcac 420
 |||
 Db 163 TGCCACACTTGTCAACATTAAGCCCGACACGCTCTGCTACTGAGCTGTGTAGGGCAC 222

QY 421 attggcgagctgtgcgt 439

Db 223 ATTTGGCAGCTGTGCCGT 241
|||||
RESULT 37
Bg258460 891 bp mRNA EST 13-FEB-2001
LOCUS Bg336190
DEFINITION 60237938F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4510616 5',
mRNA sequence.
ACCESSION Bg258460
VERSION Bg258460.1 GI:12768189
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 891)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM10392 row: 1 column: 09
High quality sequence stop: 710.
Location/Qualifiers
1. 891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4510616"
/clone_1lb="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Testis; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; Oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT 203 a 246 c 271 g 171 t
ORIGIN
Query Match 19.7%; Score 197; DB 175; Length 891;
Best Local Similarity 100.0%; Pred. No. 1.7e-93;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcgcagcacaagaattcaaacccagctcaacatcatcaccgcggacatctcccccgt 60
|||||
Db 367 TCGAGCGCAACAAGATCAAAACCCAGCTCAACCTCATCCACCGGACATCTCCCGCGCT 426
|||||
QY 61 caccagcttcgcgtgtaagaagagggcccccacccctcagtggtcccatggttcaggatga 120
|||||
Db 427 CACCAGTTCCGCTGTAAAGAGGAGGCCCCACCCCTCAGTGTGCCATGTTCAGGGGTGA 486
|||||
QY 121 atgcctcctaagtaacagctccgctccagagagggagtgagagggatgcatattac 180
|||||
Db 487 ATGCTCTCTAAGTACCAAGCTCCCTCCAGAGGAGGTGGCAGAGGATGCCATTATTAC 546
|||||
QY 181 ttgcaatcctgaagaat 197
|||||
Db 547 TTGCAATCTGAGGAAT 563
|||||
RESULT 38
Bg336190 906 bp mRNA EST 27-FEB-2001
LOCUS Bg336190
DEFINITION 602404980F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4542594 5',

mRNA sequence.
ACCESSION Bg336190
VERSION Bg336190.1 GI:13142628
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 906)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM1222 row: P column: 19
High quality sequence stop: 610.
Location/Qualifiers
1. 906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4542594"
/clone_1lb="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: POTB; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGCAG(5). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 230 a 230 c 302 g 144 t
ORIGIN
Query Match 19.5%; Score 195; DB 152; Length 906;
Best Local Similarity 100.0%; Pred. No. 1.9e-92;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 764 ggaagattcagaccctgtgtgtcggaactgcaagatgctgtgtgtgtgtgtgtgt 823
|||||
Db 32 GGAAGAGTTTACAGCTGTCTGTGTGCGGCACTGCAAGCATGCTTTGGCTGTGCGTGT 91
|||||
QY 824 gcacacctgtgtcggaagtgtctatctccggggaacacatgcccgtgagagcttgt 883
|||||
Db 92 GCACACCTGTGTGTGAAGGTGTCTATTCCGGGACACATGCCCTGCGAGGCTGTGT 151
|||||
QY 884 ccgagatggggaagaatgcccacccctcgatatacagaagccacccctggaagtggtttga 943
|||||
Db 152 CCGGATGGGAAAGATGCAACCTCTGTATATGAACCCACCTGGAAGATGTTTGA 211
|||||
QY 944 agaggaagcaagtga 958
|||||
Db 212 AGAGGAAGCAAGTGA 226
|||||
RESULT 39
Bg387104 961 bp mRNA EST 12-MAR-2001
LOCUS Bg387104
DEFINITION 602455813F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4584266 5',
mRNA sequence.
ACCESSION Bg387104
VERSION Bg387104.1 GI:13280550
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 961)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L10M1309 row: 1 column: 03
High quality sequence stop: 752.
Location/Qualifiers
1. .961
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4584266"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site:1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 228 a 269 c 262 g 202 t
ORIGIN

Query Match 19.5%; Score 195; DB 153; Length 961;
Best Local Similarity 100.0%; Pred. No. 1.9e-92;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 ccagttccgcgtgaagaaggagggcccccacccctcagtgccacagtggtcaggtgaat 122
|||||
DB 311 CCAGTTCCGCTGTAGAGAGAGAGGCCCCACCTCAGTGTGCCCATGTTCAGAGGTGAAT 370
QY 123 gctcctcaagtaaccagctccctccaggaaggagtgccagagagatgccatattactt 182
|||||
DB 371 GCCTCTCAAGTACCAAGCTCCGCCAGAGAGGAGTGCGAGAGGAGTGCATTATTACTT 430
QY 183 gcaatcctgaagaatcatagttgaaggcgtcagcttcccaactccagcagaagctgc 242
|||||
DB 431 GCAATCCTGAGGAAATCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGC 490
QY 243 aggaagtacagaagga 257
|||||
DB 491 AGGAGTACAGAGAGA 505

RESULT 40
AA346268 394 bp mRNA EST 21-APR-1997
LOCUS AA346268
DEFINITION E5752407 Greater omentum IV Homo sapiens cDNA 5' end, mRNA
sequence.
ACCESSION AA346268
VERSION AA346268.1 GI:1998525
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 394)
AUTHORS Adams M.D., Kerlavage A.R., Fleischmann R.D., Fuldner R.A., Bult

C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D., White
O., Sutton G., Blake J.A., Brenden R.C., Man-Val C., Clayton R.A.,
Cline T.R., Cotton M.D., Earle-Hughes J., Fine L.D., Fitzgerald
L.M., Fitzhugh W.M., Frilichman J.L., Geoghegan N.S., Glodok A.,
Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr., Kelley J.M.,
Kelley J.C., Liu J., Li, T., Marmaros S.M., Merick J.M.,
Moreno-Palancas R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M.,
Phillips C.A., Ryder S.E., Scott J.L., Saudak D.M., Shirley R.,
Small R.V., Springs T.A., Utterback T.R., Weidman J.F., Li Y.,
Bednarik D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,
Dimke D., Feng D.-F., Ferrie A., Fischer C., Hastings G.A., He W.M.,
Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K., Kozak D.L.,
Kunsch C., Hungjun J., Li H., Melissner P.S., Olsen H., Raymond L.,
Wei Y.F., Wing J., Xu C., Yu G.L., Ruben S.M., Dillon P.J., Fannon
M.R., Rosen C.A., Haseltine W.A., Fields C., Fraser C.M. and
Venter J.C.
TITLE Initial assessment of human gene diversity and expression patterns
JOURNAL Based upon 83 million nucleotides of cDNA sequence
MEDLINE Nature 377 (6547 Suppl), 3-174 (1995)
COMMENT 96026280
Other ESTs: TR0175624
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
1. .394
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):147767"
/db_xref="taxon:9606"
/clone_lib="Greater omentum IV"
/dev_stage="adult"
/note="Organ: omentum; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 81 a 108 c 108 g 96 t 1 others
ORIGIN

Query Match 18.5%; Score 185; DB 6; Length 394;
Best Local Similarity 100.0%; Pred. No. 3.8e-87;
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 gcaacgagatcaacacagcgcttgcgaagtatctgtcgaagagagacgcgcctggc 562
|||||
DB 85 GCACGAGATCAACACACACGCGCTTGCCAAATCTTGCTGCAGAGAGAACGCGCTTGGC 144
QY 563 atctttgggaagcgcgtcaccccttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 622
|||||
DB 145 ATCTTTGGGAAAGCCGCTTCAACCTTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 204
QY 623 gctccagcagatcccaaccagatgcaggaaggttcctgcgacacatagatattctctgc 682
|||||
DB 205 GCTCCAGCAGTACACACACACAGTGCAGAGAGTCTGCACACACATCAATGATATTCCTGC 264
QY 683 caaat 687
|||||
DB 265 CAAT 269

RESULT 41
BG116283 936 bp mRNA EST 30-JAN-2001
LOCUS BG116283
DEFINITION 602318546f1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4418843 5',
mRNA sequence.
ACCESSION BG116283
VERSION BG116283.1 GI:12609789

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 936) Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
plate: LLM10153 row: 1 column: 12
High quality sequence stop: 699.

FEATURES
source
1. 936
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4418843"
/clone_1lb="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-Sport6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 209 a 267 c 265 g 195 t

ORIGIN

Query Match 18.5%; Score 185; DB 174; Length 936;
Best Local Similarity 100.0%; Pred. No. 4.1e-87;
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 ctacaccacgcagacatctcccccgtcaccagcttcgcgtgtaagaaggagccccc 91
|||||
Db 390 CTATCACCACCGGACATCTCCCTCTCACCAGTTTCCGCTTAGAAGAGAGGCCCC 449

QY 92 accctcagtgcccatggttcagggtgaatgcctcctcctcaagtcacagctcgtccagg 151
|||||
Db 450 ACCCTCAGTGTGCCCATGTTCAAGGTGAATGCTCTCTCAAGTACCAAGCTCCGTCACAG 509

QY 152 agggagtgagagagagagatgcatattctgcatcctcagagatcattcatggtgagcg 211
|||||
Db 510 AGGAGTGGCAGAGGAGTGCATTACTTCAATCTTGAGGAATCATAGTTGAGCG 569

QY 212 ctgca 216
|||||
Db 570 CTGCA 574

RESULT 42
BF686235 547 bp mRNA EST 22-DEC-2000
LOCUS 602143667F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4304829 5',
DEFINITION mRNA sequence.
ACCESSION BF686235
VERSION BF686235.1 GI:11971643
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 547) Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
plate: LLM1170 row: m column: 22
High quality sequence stop: 545.

FEATURES
source
1. 547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4304829"
/clone_1lb="NIH_MGC_46"
/tissue_type="leiomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(5). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 127 a 151 c 149 g 120 t

ORIGIN

Query Match 18.2%; Score 182; DB 168; Length 547;
Best Local Similarity 100.0%; Pred. No. 1.5e-85;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgagcgcacaaagttcaaacccagctcaactatcaccacccgacatctcccccgtc 60
|||||
Db 366 TCGAGCGCACAAAGATTCAAAACCCAGCTCAACCTCATCCACCGGACATCTTCCCTGCT 425

QY 61 caccagttccgcgtgaagaaggagggccaccctcagtgcccatggttcaggggtga 120
|||||
Db 426 CACCAAGTTCCGCTGTAGAAGAGAGAGGCCCAACCTCATAGTGTGCCAAGTTCAAGGGGA 485

QY 121 atgcctcctcaagtaaccaagctccgtcccaaggagagtgagagagaggtccattatc 180
|||||
Db 486 ATGCCTCCTCAAGTACCAAGCTCCGTCACAGAGGAGTGCAGAGGATGCATTATTAC 545

QY 181 tt 182
||
Db 546 TT 547

RESULT 43
BE794311 688 bp mRNA EST 20-SEP-2000
LOCUS 601591442F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945538 5',
DEFINITION mRNA sequence.
ACCESSION BE794311
VERSION BE794311.1 GI:10215496
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 688) Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DTF

URES	Location/Qualifiers
source	1. .688

```

source
1. .852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4133442"
/clone_1lb="NH_MGC_54"
/tissue_type="from chronic myelogenous leukemia"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: bone marrow; Vector: pDMR-LIB (Clontech);
Site.1: SfiI (ggccgcctggcc); Site.2: SfiI (ggccatattggc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCCGAGCGCCGACATG-drr(30)BN-3'
(Where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT      188 a      220 c      270 g      174 t
ORIGIN

```

	LOCUS	DEFINITION	441 bp	mrna	EST	20-JUN-1995
Qy	H03317	y147e10.r1 Soares placenta		Nb2HP Homo sapiens	cDNA clone	
Db	103	TCAAGCCTGGCTCCAGAGTACACAAACAGTCCAGAGAGTCTCTACACATCAGTA	163			
Qy	673	tgattccctgcacaatggccttcagaaggggctgagatccagtcctcagtggaagaat	722			
Db	163	TGATTCCTCGCCAAATGCTTCAGGAAAGGGGCTGAGATCTCCAGTCTCGAGTGGAAAGAT	222			
Qy	733	tgatcagttgcctcttttggaacatgtaatttggagaaggttcagacctgtcgtgt	787			
Db	223	TGATCAGTTCCTGTGTGGACATGTATTTGGAAAGATTTCAGACCTGTGCTGTGT	277			
RESULT	45					

LOCUS	H03317	441 bp	mRNA	EST	20-JUN-1995
DEFINITION	Y47610.i.1 Soares placenta Nb3HP Homo sapiens cDNA clone				
ACCESSION	IMAGE:151914 5', mRNA sequence.				
VERSION	H03317				
KEYWORDS	H03317.1	GI:866250			
SOURCE	EST.				
ORGANISM	human.				
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 441)				
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucab,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rikkin,L., Rohlfing,T., Soares,M., Tan,F., TrevasKis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.				
TITLE	The Washu-Merck EST project				
JOURNAL	Unpublished (1995)				
COMMENT	Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800				

High quality sequence stops: 317
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1152 Std Error: 0.00
 Seq primer: M13Rpl
 High quality sequence stop: 317.
 Location/Qualifiers

```
source
1. .441
/organism="Homo sapiens"
/db_xref="GDB:564064"
/db_xref="taxon:9606"
/clone="IMAGE:151914"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; vector: pTV73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGAGAGAAATCCGCGCGAGCAATTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTV73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo. "
```

BASE COUNT 96 a 102 c 134 g 102 t 7 others

ORIGIN

```
Query Match 17.48; Score 174; DB 157; Length 441;
Best Local Similarity 100.0%; Pred. No. 2.8e-81;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 757 gtgattggaagagttccaacctgtctgtgctgcaactgcaagcatgctgtgtg 816
|||||
Db 55 gtagtttggaaagtccaacctgtctgtgctgcaactgcaagcatgctgtgtg 114
|||||

QY 817 cgtgtgtcacacctgtgctgtgaaagtgtctatccggggaaacacatgacctgag 876
|||||
Db 115 CGCTGCTGCACACCTCTGCTGCAAGTGTCTATTCGGGGACACCATGCTTGCAGG 174
|||||

QY 877 ctctgtcggatgaggaaatgcaacctcctctatatacatgaagccacctg 930
|||||
Db 175 CTCTGCTCCGATGGGAAAGATGCCACCTCTGATACATGAAGCACCCCTGG 228
|||||

Search completed: October 28, 2001, 22:09:06
Job time: 2744 sec

